INTERLOCUTOR EFFECTS ON SOCIOLINGUISTIC VARIATION IN L2 FRENCH

Mark Alan Black

Submitted to the faculty of the University Graduate School

in partial fulfillment of the requirements

for the degree

Doctor of Philosophy

in the Department of French and Italian,

Indiana University

February 2021

Accepted by the Graduate Faculty, Indiana University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Doctoral Committee:

_____

Committee Chairperson, Julie Auger, PhD

_____

Committee Chairperson, Laurent Dekydtspotter, PhD

_____

Committee Member, Barbara Vance, PhD

_____

Committee Member, Kelly Sax, PhD

_____

Committee Member, Kimberly Geeslin, PhD

Date of Defense: July 16th, 2020

For Amber and Leona

Acknowledgements

I begin by thanking the co-directors of my dissertation, Julie Auger and Laurent Dekydtspotter. I could not have completed this project without their steady guidance and extreme patience, from conception of the topic to the final edits. I am also very grateful for the members of my dissertation committee beyond their roles in the dissertation-writing process: Kelly Sax, who guided me through my graduate studies as an excellent instructor, mentor, and supervisor; Barbara Vance, for her guidance as an instructor and for her advice during my job search preparation; and Kim Geeslin, whose presence as outside committee member became more and more invaluable as my project became more and more aligned with aspects of her and her students' research programs.

In the French and Italian department, I am also very grateful to Kevin Rottet for his meticulously prepared courses in French Linguistics and for his comments on my presentations of research at various conferences in Bloomington. I would also like to thank Albert Valdman, not only for his comments on my research presentations at IU, but also because his textbook was my very first introduction to the French language as a high school student in 1994.

I am also indebted to other scholars who visited Bloomington during my time there. Luke Plonsky was gracious in sharing with me some of his unpublished research (with Laura Gurzynski-Weiss) on interlocutor individual differences. Many thanks to Aiden Coveney, who took time from his Bloomington visit to discuss his research with me and to offer suggestions on my own research.

I also owe great thanks to Charlène Gilbert and Arielle Roadman for their help in coding the variationist analyses, and also to my officemates Kelly Kasper-Cushman and Laura Demsey, who provided a sounding board for various aspects of the dissertation as well as for the job search. Thanks also to Carly Bahler for providing last-minute editing and proofreading support. And finally, many thanks to Rodica Frimu, an extraordinary friend to Amber and me during our

challenging coursework and research projects together in Bloomington, our summers with the IUHPFL in Brittany and in Paris, our dissertation tribulations in Lille, and our adventures in Brussels and Knoxville by way of Nebraska.

This dissertation would also not have been possible without the generous offering of time and resources from many friends and colleagues in France. From the Pau community, I would like to thank Soubha Essafi, Chloé Butler, Freddie Puszka, Thomas Dunning-Laredo, Marc Artzrouni, and Aaron Murray-Nellis for their help in organizing my project and recruiting participants. I also wish to thank Amanda Edmonds for providing me with lodging upon my arrival in Pau, for recruiting participants, and for providing me with the opportunity to present my research to the scholars at the Université de Pau. Amanda also coordinated Bryan Donaldson's visit to Pau; the discussions I had with Amanda and Bryan during his visit were instrumental in the organization of my project. A second discussion I had with Bryan in Bloomington in 2019 was immensely helpful for my statistical analysis. In Lille, I wish to thank Sarah Moreau and Capucine Fosse for their enormous help in organizing my project there. Thanks also to Rebecca Grey, Jeremy Levin, Bailey Wamp, Kevin Bolle, Chad Langford, and all colleagues at Université Lille 3.

Finally, I thank my family for their unending patience and support: my wife Amber and her family, my daughter Leona, my parents Michael and Cynthia, my brother Greg, my sister Becky and her family, my aunt Nancy, and my grandmother Phyllis.

The development of this project and interpretation of the data therein were aided by valuable discussion with audiences at the AILA Symposium on Interlocutor Individual Differences (Indiana University, October 2015), SLRF 36 (The Ohio State University, October 2017), NWAV 47 (New York University, October 2018), and SLRF 38 (Michigan State University, September 2019).

INTERLOCUTOR EFFECTS ON SOCIOLINGUISTIC VARIATION IN L2 FRENCH

Mark Alan Black

The ability to speak in a second language (L2) requires a certain level of linguistic proficiency, but the ability to *live* in a second language requires a certain level of *sociolinguistic* proficiency. L2 sociolinguistic variables present acquisitional challenges for language learners, since informal discourse features are largely absent from classroom-based input but frequent in native speakers' informal communication.

In this dissertation, I examine how L2 sociolinguistic performance can be influenced by a specific social characteristic: the interlocutor's native language status vis-à-vis the language of communication. That is, how does learner speech change in conversation with a native speaker compared to conversation with another learner who shares the same L1? While previous studies have examined this interlocutor characteristic on measures of grammatical proficiency in classroom-based learners, few studies have measured its effect on sociolinguistic performance, especially in highly advanced learners. My data focus on two sociolinguistic features that frequently appear in informal French: *ne*-deletion (ND) and subject doubling (SD).

I examine the interlocutor effect on these variables in two groups of learners: study-abroad students at low-advanced proficiency and highly proficient near-native speakers. Both groups were recorded in informal one-on-one conversations with a native and non-native French interlocutor. Study-abroad students demonstrated significantly higher rates of ND and SD (characteristic of more informal, nativelike speech) in conversation with a native French speaker than when speaking with another study-abroad student. Furthermore, a variationist analysis revealed interlocutor language status as the most significant social factor influencing variation for ND and SD. In near-native speakers, only marginal differences in ND and SD frequency were detected

across interlocutor language statuses, suggesting a diminishing influence as proficiency increases. The results demonstrate that researchers must be aware of this interlocutor effect when designing tasks that evaluate sociolinguistic performance in learners.

Table of Contents

Chapter 1: Introduction 1
Chapter 2: Review of the literature: SLA, sociolinguistic variation, and interlocutor effects.....5
2.1 Second language acquisition outcomes 5
2.1.1 The syntax-discourse interface **..... 7**
2.2 Factors influencing sociolinguistic variation 8
2.3 A definition of "interlocutor" 11
2.4 Accommodation Theory and Audience Design 12
2.4.1 Accommodation Theory in SLA 14
2.4.2 Priming effects on convergence 16
2.5 Interlocutor language background influencing native speakers 17
2.6 Interlocutor language background and SLA 18
2.7 Sociolinguistic variables and study abroad 24
2.8 Type I and Type II variation 25
2.9 Sociolinguistic variation in L2 French 32
2.9.1 Sociolinguistic variation in L2 French: The question of interlocutor L1 effects 38
2.10 Variationist approaches to SLA 45
2.11 Summary of Interlocutor effects, SLA, and sociolinguistic variation 48
Chapter 3: Methodology 50
3.1 Informal discourse 50
3.2 Research questions 55
3.3 Hypotheses 57
3.4 Pilot study 57
3.4.1 Summary of pilot study and results 58
3.4.2 From the pilot to the current study 60
3.4.3 Informing the methodology of the current study 61
3.5 The current study 68
3.5.1 SA learner group 68
3.5.2 Near-NS groups 73
3.6 Corpus 88
3.6.1 Quantitative dominance 90
3.6.2 Informal features: Near-NSs 98
3.6.3 Informal features: SA learners 105
3.6.4 Which sociolinguistic variables to analyze? 108
3.6.5 Statistical tools for variationist analysis 110
3.7 Summary of methodology 111
Chapter 4: *Ne*-retention 113
4.1 Background on French negation 113

4.2 Studies on *ne*-retention in L1 French 116
4.2.1 Factors influencing *ne*-retention in L1 French 120
4.3 *Ne*-retention in L2 French 126
4.3.1 Factors influencing *ne*-retention in L2 French 132
4.4 Results: *ne*-retention 134
4.4.1 SA Learners: Overall results 136
4.4.2 SA Learners: *Ne*-retention by interlocutor type 140
4.4.3 Near-NSs: Overall results 145
4.4.4 Near-NSs: *Ne*-retention by interlocutor type 157
4.4.5 *Ne*-retention results for L2 French speakers: Conclusions 161
4.5 Variationist analysis: *Ne*-retention 161
4.5.1 Linguistic factors 162
4.5.2 Commentary on additional linguistic factors 166
4.5.3 Extralinguistic factors: SA learners 169
4.5.4 Extralinguistic factors: Near-NSs 170
4.5.5 Extralinguistic factors: Interlocutors 171
4.5.6 Sociostylistic factors 171
4.5.7 Results: SA learners 175
4.5.8 Results: Near-NS groups and interlocutors 180
4.5.9 Near-NS groups and interlocutors: Commentary 188
4.6 Hypotheses revisited: *Ne*-retention 203
Chapter 5: Subject doubling 205
5.1 Background on subject doubling in French 206
5.1.1 Strong pronouns and subject doubling 206
5.1.2 The framework of subject doubling analyses 209
5.1.3 Subject doubling or left-dislocation? 211
5.1.4 Subject doubling and prescriptivism 222
5.2 Studies on subject doubling in L1 French 226
5.2.1 Subject doubling in children 228
5.2.2 Non-Hexagonal French 230
5.2.3 Factors influencing subject doubling in L1 French 232
5.3 Subject doubling in L2 French 244
5.3.1 Subject doubling in the language learning classroom 253
5.3.2 Subject doubling and pedagogical norms 256
5.4 Results: subject doubling 260
5.4.1 Native speaker judgments on clitic types 260
5.4.2 Inclusions and exclusions 264
5.4.3 SA learners: Overall results 275
5.4.4 SA learners: Subject doubling by interlocutor type 280
5.4.5 Near-NSs: Overall results 283

5.4.6 Near-NSs: Subject doubling by interlocutor type ... 293
5.4.7 Strong pronouns ... 297
5.4.8 Personal vs. neuter pronouns ... 298
5.5 Variationist analysis: Subject doubling ... 307
5.5.1 Linguistic factors ... 307
5.5.2 Extralinguistic factors ... 314
5.5.3 Sociostylistic factors ... 314
5.5.4 Results: SA learners ... 315
5.5.5 Results: Near-NS group and interlocutors ... 321
5.5.6 Summary of variationist analyses on subject doubling ... 337
5.6 Hypotheses revisited: subject doubling ... 338
Chapter 6: *Ne*-retention/subject doubling interactions and conclusions ... 340
6.1 L2 speakers of French and the interlocutor effect: A summary ... 340
6.2 *Ne*-retention use in SD contexts ... 346
6.3 Correlations of *ne*-deletion and SD: Group analyses ... 353
6.3.1 Correlations of *ne*-deletion and SD: SA learners ... 355
6.3.2 Correlations of *ne*-deletion and SD: Near-NSs ... 358
6.3.3 Correlations of *ne*-deletion and SD: Interlocutors ... 360
6.3.4 *Ne*-deletion and SD: Comparisons with Coveney (2005) ... 362
6.3.5 Correlations of *ne*-deletion and SD (and beyond): Future considerations ... 364
6.4 Conclusions from the current study: Pedagogical implications ... 365
6.5 Conclusions from the current study: Methodological implications ... 369
6.5.1 Limitations in the current study ... 369
6.6 Revisiting a language learner's questions ... 371
Bibliography ... 373
Appendix A: Pilot study ... 413
A.1 Participants ... 413
A.2 Procedure ... 415
A.3 Results: Acceptability Judgment Task and language security index ... 416
A.4 Results: *Ne*-retention ... 419
A.5 Pilot study variationist analysis ... 423
A.6 Limitations in the pilot study ... 426
Appendix B: Pragmatic particles in informal discourse ... 428
B.1 Pragmatic particles *quoi*, *genre*, *machin*, *bref* ... 428
B.2 Distribution of pragmatic particles in speaker groups ... 430
Appendix C: Written topic prompts for SA learners ... 433
Appendix D: Proficiency measure (*c*-test) ... 434
Appendix E: Language background questionnaire ... 435
Appendix F: Acceptability Judgment Task for near-native speakers ... 441
Appendix G: Non-significant factor groups in variationist analyses ... 448
G.1. Non-significant factor groups for *ne*-retention ... 448

G.2. Non-significant factor groups for subject doubling 454
CURRICULUM VITAE

Chapter 1: Introduction

This dissertation addresses a gap in the research on interlocutor effects in the intersection of sociolinguistics and second language acquisition. Through an empirical examination of sociolinguistic variation in English-speaking learners of French, I provide evidence for an effect of the interlocutor's language background (as a native or non-native speaker) conditioning the spontaneous oral production of two sociolinguistic variables in French for certain groups of learners.

The origins of this contribution lie in informal observations that I made as a language student working to gain proficiency in French as a foreign language. During my sojourns abroad, I found myself paying particular attention to the ways in which Anglophones (including myself) interacted in the presence of native French speakers, reflecting on the choice of vocabulary as well as grammatical structures in French when speaking with native speakers compared with non-native speakers. For example, an English-speaking friend would sometimes use *ma caisse* ('my car,' informal; cf. 'my ride') to refer to a car when addressing his French friends, then switch to *ma voiture* ('my car') when addressing other Anglophones, including me, in the same conversation. Why did my friend use different words with different speakers? Regarding my own language abilities as an adult learner of French, why did I feel like my French "flowed" more naturally with native speakers but often required more of an effort with non-native speakers, even those who were already fluent in French? Was this the case for other learners of French?

My goal, while living in France, was to master the language to the best of my abilities, so that I might be able to pass as a native speaker, however briefly, in a given interaction. Though at that time I did not conceive of it in terms of sociolinguistic performance, I knew that "mastery" of the language, at least in my definition, included more than simply using the correct verb

conjugations, syntax, and vocabulary. However, I sometimes felt that the speaking context in which I found myself did not allow me to show that I had mastered, at least to a certain extent, the sociolinguistic aspects of the language. When speaking casually in French with other native English speakers, the choice of French sometimes seemed artificial, and I found myself consciously producing certain forms that reflected this artificiality, such as maintenance of standard bipartite negation as *ne…pas* rather than dropping *ne*, as is done in casual French, or articulation of *je suis* ('I am') as separate words rather than informal contraction to *j'suis* or *chui*. Sometimes, these choices were made in order to accommodate my interlocutor's lower proficiency level; other times, I felt it necessary to assert my identity as an American, and my French prosody became less nativelike; at still other times, it became simpler to just switch into English. I began, then, questioning to what extent other learners of French made conscious (or subconscious) reflections on language choices such as these.

After returning to the United States, I took the OPI (Oral Proficiency Interview) administered by the ACTFL (American Council on the Teaching of Foreign Languages) as part of an assessment of my competence in French for my studies in language education. At a certain point during the interview, which was conducted all in French, I became aware that my interviewer was a native speaker of English, which caused me to more consciously monitor certain aspects of my speech, rendering this interaction even more "artificial" than an interaction with a native French speaker evaluating my speaking ability. I thus felt that my ability to "perform" as nativelike as possible was compromised by the status of my interlocutor as a non-native speaker.

Based on these observations, I became interested in the obstacles that learners might face in preventing them from demonstrating an optimal (and, potentially, nativelike) level of performance in a given social interaction, especially those obstacles that can be manipulated within

the parameters of a controlled oral production task (e.g., requiring the learner to speak entirely in the target language, or pairing two speakers from different social backgrounds). On a broader, societal scale—that is, in a given language community—a speaker may consciously choose to interact with, or avoid, certain interlocutors (such as other native English speakers); in a given social situation, two speakers proficient in more than one language may choose one language over another (such as English when both interlocutors are native speakers of this language). However, in an empirical study, if I could control the language of communication and the choice of interlocutor—in other words, if a speaker must speak French with a given conversation partner—would there be any detectable differences in this speaker's use of the language across different interlocutors? This project is thus an attempt to analyze such behavior in order to better understand how language learners develop and demonstrate sociolinguistic ability, and how researchers in second language acquisition can inform their methodological decisions in creating oral production tasks for learners.

My initial observations, and the current project stemming from these observations, fundamentally involve two broad areas of language study: second language acquisition (and its outcomes) and sociolinguistic variation (including social factors such as interlocutor effects). The dissertation aims to answer some basic questions regarding the role of the interlocutor's language background against the backdrop of these two broad fields; crucially, such an effect has been rarely studied in the intersection of these two contexts. The data obtained to test this interlocutor effect consist of recorded spontaneous conversations involving, along with native French speakers, learners with a similar language profile as mine: adult native English speakers living (at least temporarily) in France and learning French as a second language.

In Chapter 2, I introduce the background literature on relevant aspects of second language acquisition, sociolinguistic variation, and interlocutor effects for this dissertation. In Chapter 3, I outline the methodology of the current empirical study involving the collection of a new corpus of learner and native speaker oral production data in French, consisting of learner groups at two proficiency levels (pre-advanced and near-native); I also outline a methodology testing the relevance of learner perceptions of the interlocutor's native language background. Subsequently, I select two specific sociolinguistic variables in order to analyze the extent of this interlocutor effect on sociolinguistic performance, detailed in Chapter 4 for the variable of *ne*-retention and Chapter 5 for the variable of subject doubling. In this dissertation, in addition to addressing a dearth of learner production data on the latter variable, I use a variationist analysis to provide empirical evidence for this interlocutor effect conditioning *both* variables in the pre-advanced learner group, but not in the near-native group. In the sixth and final chapter, I examine the interaction of these two variables and detail the pedagogical and methodological implications drawn from the results of this project.

Chapter 2: Review of the literature: SLA, sociolinguistic variation, and interlocutor effects

In this chapter, I provide a background on the existing literature relevant to the intersections of second language acquisition, sociolinguistic variation, and interlocutor effects. I begin with a review of second language acquisition concepts relevant to the current study, followed by a discussion of factors influencing sociolinguistic variation in all speakers. From these factors, I then focus on factors relating to the interlocutor, involving a discussion of theories on convergence and divergence strategies between speaker and interlocutor. I continue with an examination of the applications of interlocutor effects within a broader discussion of sociolinguistic variation research on learners before focusing on an overview of studies in this domain for learners of French. Finally, I motivate a variationist analysis for the statistical detection of an interlocutor effect.

2.1 Second language acquisition outcomes

Though much of second language acquisition (SLA) research focuses on the stages and processes in acquiring language, considerable attention has been devoted to final outcomes for adult second language (L2) learners (e.g., Birdsong, 1992; Coppieters, 1987; Lardiere, 2007). Much of this research has concerned ultimate attainment in terms of syntax/morphology and phonology. Sociolinguistic development has received comparatively less attention, even though early SLA research in this domain identified sociolinguistic competence as integral, along with grammatical competence, in developing overall communicative competence (Canale & Swain, 1980). Geeslin and Long (2014: 78) have recently commented that "studies that focus on the connection between models of learner language, acquisition, and use, on the one hand, and the effects of various social factors, on the other hand, are quite scarce." The intersection of sociolinguistics and second language acquisition can nevertheless be seen in a recent proliferation

of sociolinguistics handbooks containing chapters on SLA (e.g., Geeslin, 2011) and SLA handbooks containing chapters on social factors (e.g., Gass, Behney, & Plonsky, 2013; Hummel, 2014; VanPatten & Williams, 2014) and sociolinguistic variation (e.g., Bayley & Tarone, 2011; Gudmestad, 2014; Kanwit, 2018), while some textbooks focus specifically on this intersection (e.g., Adamson, 2009; Geeslin & Long, 2014).

The effects of social factors are indeed relevant in developing sociolinguistic competence in both native (L1) and non-native speakers as they learn which speech patterns are sociolinguistically and pragmatically acceptable (and preferred) in each social setting; that is, to paraphrase Fishman's (1965: 67) famous observation, they learn "who says what to whom, where, and when." A more recent description by Geeslin (2019) states that not all native speakers speak the same way, and they understand that different ways of speaking are necessary in order to "fit in" with a given speaker group in a given context. To successfully interact with such groups, L2 learners must acquire this sociolinguistic competence as well. Geeslin and Long (2014) succinctly summarize these challenges as follows: the ability to speak in a second language requires a certain level of linguistic proficiency, but the ability to *live* in a second language requires a certain level of *sociolinguistic* proficiency. As with grammatical proficiency, sociolinguistic proficiency has been shown to be influenced by multiple factors, such as age of initial exposure to the target language as well as the depth of the learner's integration into the target language community. Furthermore, L2 pragmatic behaviors may remain entrenched in L1 patterns, so that L2 sociolinguistic ability "lags" behind L2 grammatical development; importantly, sociopragmatic development does not necessarily follow from grammatical development (Bardovi-Harlig, 1999, 2012, 2013). Ultimately, learners may never reach nativelike sociolinguistic usage patterns, and learner individual differences (Dörnyei, 2005) may play a role in final learning outcomes.

2.1.1 The syntax-discourse interface

Sociolinguistic ability as measured in learners has also been posited to be constrained by residual non-nativelike "instability" between the various cognitive domains, known as the Interface Hypothesis (Sorace, 2006, 2011). According to recent versions of this hypothesis (cf., e.g., Sorace & Serratrice, 2009; Tsimpli & Sorace, 2006; White, 2009), nativelike syntax in learners is theoretically perfectly acquirable where purely linguistic, or internal, domains are involved, such as syntax-semantics or phonology-morphology interfaces. External interfaces, such as syntax-pragmatics, or syntax-discourse, are posited to be unstable—that is, subject to permanent optionality in learners. Tsimpli and Sorace (2006) argue that, for example, syntactic parameters can be reset in the L2, but discursive features are not consistently reset. Since sociolinguistic ability concerns pragmatic and discourse domains, linguistic structures involving, for example, syntactic or morphosyntactic variation in specific discursive contexts would be predicted to pose an acquisitional challenge for learners. The inability to reset non-linguistic parameters may be a structural difference in the interfaces of learners, or it may simply manifest in on-line processing delays in learners (Sorace, 2006).

Other recent investigations into the acquisition process have challenged the Interface Hypothesis on several fronts. Mapping properties at certain internal interfaces (e.g., syntax-semantics) has been shown to be an easier acquisitional challenge than at other internal interfaces (Slabakova, 2006); furthermore, mapping properties to external interfaces may not necessarily be more difficult than at internal interfaces (e.g., Montrul, 2011; Slabakova, 2014). A body of research across multiple languages (Donaldson, 2008, 2011a, 2011b; Leal Mendez, Rothman, & Slabakova, 2015; Slabakova, Kempchinsky, & Rothman, 2012) has likewise advanced evidence of successful acquisition at external interfaces. At high levels of proficiency, the challenge of

nativelike acquisition may not in fact concern interfaces, but rather functional morphology (Slabakova, 2014, 2019). In any case, for certain learners at either an end state or steady state of L2 acquisition, the upper limits of what is cognitively possible when measuring sociolinguistic ability may not be restricted by fundamental differences in learner grammars. Furthermore, an empirical examination of L2 sociolinguistic ability would benefit from the inclusion of highly proficient learners with potential successful acquisition of internal and external domains.

2.2 Factors influencing sociolinguistic variation

Let us now move from what is cognitively possible in learner grammars to how situational factors may constrain measures of learner performance. Based on a body of literature examining sociolinguistic variation in L1 and L2 speakers, particularly the early work of Tarone (1979), Geeslin and Long (2014) identify three aspects of the social interaction setting that affect language behavior. These categories include factors related to the speaker, factors related to the interlocutor, and factors related to the speech context (see Table 2-1 on p. 10). Factors related to the speaker have received the most attention in the sociolinguistics literature, and they can be divided into two categories: 1) objective characteristics, including age, ethnicity, socioeconomic status, occupation, education, country of origin, and others (Geeslin & Long, 2014: 149); and 2) behavioral characteristics, such as personality traits, ability or aptitude, motivation, learning styles, and language learning strategies (Dörnyei, 2005). Though both categories can serve to identify one speaker as different from another, the latter category is more commonly associated with *individual differences*, to which Dörnyei (2005: 4) applies the definition of "dimensions of enduring personal characteristics that are assumed to apply to everybody and on which people differ by degree."

Factors related to the interlocutor have included age, gender, ethnicity, and relationship to speaker. These factors can also be called *interlocutor effects* or *interlocutor individual differences*.

Note that the "relationship" factor may be bi-directional; that is, my relationship with my interlocutor often reflects the interlocutor's relationship with me (such as friend-friend or stranger-stranger). This relationship can reflect real or perceived social and psychological distance, often conditioned by power balances that can be symmetrical (such as friend-friend or certain coworker-coworker relationships) or asymmetrical (such as teacher-student or boss-employee relationships).

Finally, factors related to the speech context have included topic, task, and discourse type. These factors have been involved in sociolinguistic analysis since its early days. In an attempt to minimize the "observer's paradox," Labov (1972), in his sociolinguistic interviews, included topics of conversation that were likely to elicit emotional or personal reactions from speakers (such as "danger of death" questions), so that the speakers would focus more on the content of their speech rather than on the form, producing more naturalistic output. Concerning task, Labov also had his speakers read a prose passage as well as lists of isolated words and minimal pairs, with a more "careful" style of speech characteristic of each successive reading task compared with the interview task. Discourse type, which can be closely dependent on task, concerns the different discursive structures characteristic of speech, such as a narrative, a dialogue, or a role play. Other elements of the discourse setting (such as the presence of the researcher, or the location of recording equipment) can also influence the output obtained from speakers (cf. Tarone, 1979). For each of the factors related to the speech context, the speaker's goals or expectations may be manipulated, and differences in the features of the speaker's speech may be observed.

Table 2-1. Examples of social factors examined in variationist sociolinguistic studies (reproduced from Geeslin & Long, 2014: 151)

Related to the speaker	**Related to the interlocutor**	**Related to the speech context**
Age	Age	Topic
Gender	Gender	Task
Ethnicity	Ethnicity	Discourse type
Relationship to interlocutor (family, friend, stranger, etc.)	Relationship to speaker (family, friend, stranger, etc.)	
Education		
Occupation		
Income level		
Social class		
Geographic location		
Country of origin		

For variation in L2 speakers specifically, many studies have identified individual factors from this list as producing significant effects. For example, task-based variability was found in Larsen-Freeman's study (1975) of grammatical morphemes of ESL students. In Adamson and Regan (1991), a significant difference in the L2 production of the English *-ing* morpheme was found to be due not only to linguistic constraints but also to speaker gender, which was more important than the effect of style; male speakers tended to increase their use of the informal variant even in formal styles, due to the prestige of this informal variant associated with masculinity. Young (1988) found that the ethnicity of the interlocutor had a significant effect on the production of obligatory /s/-plural marking in Chinese ESL speakers, with higher production of /s/-plural marking associated with high social convergence between the speaker and interlocutor, at least for learners with advanced proficiency.

It is important to note that, as Bayley (2005) points out, early research on variation in SLA (e.g., Beebe, 1977; Ellis, 1987; Tarone, 1985) attempted to explain learners' linguistic choices by reference to a single co-occurring contextual factor; more recent studies have proceeded from the

notion that multiple factors (both linguistic and extralinguistic) influence the choice of one variant over another. All subsequent discussion will be thus be treated with this caveat in mind; more details on this distinction will be provided in section 2.10.

2.3 A definition of "interlocutor"

With evidence that interlocutor characteristics can influence variation (e.g., Young, 1988), let us first establish what is meant by "interlocutor" before addressing its potential effects on the language habits of L1 and L2 speakers. Geeslin (2020) cites work from the SLA literature (Gass, Mackey, & Pica, 1998; Long, 1996; Mackey, 2012; VanPatten & Benati, 2010) in which the role of the interlocutor is not clearly defined but in which the interlocutor is relevant when measuring certain aspects of language acquisition, such as the role of feedback provided by the interlocutor (Gurzynski-Weiss, 2010). In previous studies more broadly, "interlocutor" has often been interchangeable with "interviewer," as is the case in many studies involving sociolinguistic interviews (e.g., Labov, 1972); in SLA literature, such a role almost always involves an interlocutor with more knowledge of the target language and thus often "in control of the communicative exchange" (Gurzynski-Weiss, 2013: 531).

Geeslin (2020: 136) defines the interlocutor as "the input provider and the communicative partner of the learner." Though the role of "input provider" may connote an unequal relationship between speakers, especially in the SLA context, Geeslin's definition does not presuppose, for example, an interviewer-interviewee relationship or an inherent difference in language proficiency between speakers. Furthermore, while the role of the input provided by the interlocutor may be less critical in studies of L1 speakers, and the amount and nature of such input may vary widely depending on the situation under which speakers are evaluated, such input is nevertheless influential in the choices all speakers make, as we will see in the review of Accommodation Theory

and Audience Design. With a view toward the crucial role of the social and linguistic characteristics of the interlocutor in measures of second language acquisition in the current study, I adopt Geeslin's definition while emphasizing that linguistic input from the interlocutor is necessarily accompanied by the interlocutor's social characteristics, both of which, in a given communicative context, define relative social proximity to, or distance from, the speaker (viz. language learner) under study.

2.4 Accommodation Theory and Audience Design

Since much of sociolinguistics research is concerned with how speakers use language in specific communicative situations, measures of sociolinguistic behavior involving spontaneous, real-time, oral production typically examine a speaker interacting with at least one other speaker in the conversational sphere. Fundamentally, and for various reasons, speakers often change their speech according to the person to whom they are speaking. The potential impact of the various individual differences that the interlocutor brings to each conversation is thus relevant when interpreting language behavior in individual speakers. In this section, I outline the contributions made by Accommodation Theory and Audience Design regarding interlocutor effects on language variation, including the application of these theories to measures of second language acquisition.

Accommodation Theory (Giles & Powesland, 1975) has been a highly influential approach in analyzing variation in individual speakers. In this theory, speakers may change their speaking style to either converge with or diverge from their interlocutor's speaking style, or they may maintain a speaking style that is not influenced by their interlocutor's speech. Convergence tends to occur when the speaker and her interlocutor are on equal terms—that is, when the social/psychological distance between a speaker and her interlocutor is small—and the speaker consciously or subconsciously adopts features of her interlocutor's speech (or features that she

believes are stereotypes of her interlocutor's speech; cf. Thakerer, Giles, & Cheshire, 1982), in order to highlight proximity with her interlocutor and thus facilitate a positive conversational interaction. On a fundamental psychological level, the idea is that the more similar we seem to our interlocutor, the more our interlocutor will like and/or respect us. Divergence occurs when the speaker wishes to emphasize distinctiveness from her interlocutor or to establish or reaffirm her own identity, such as when the social/psychological distance between speaker and interlocutor is large (e.g., in an imbalance of power), by avoiding certain features of her interlocutor's speech or by producing features that are not part of her interlocutor's speech patterns. In-group versus out-group distinctions have been shown to be maintained by divergence strategies (Bourhis, Giles, Leyens, & Tajfel, 1979; Flikeid & Péronnet, 1989).

The *Audience Design* theory (Bell, 1984) expands on the concept of convergence from Accommodation Theory as an explanation for *style-shifting*, where the speaker uses more formal or informal features (or styles) of speech according to the topic or setting, which is in turn associated with a different audience (or "classes of persons"). This audience may include not only the interlocutor (the addressee) but other listeners who are not directly addressed, such as in Bell's analysis of radio broadcasters. While research on interlocutor effects is often situated in either Accommodation Theory or Audience Design, these two approaches do not necessarily represent conflicting views, but rather different ways of explaining how speakers modify their speech in a given social situation. For example, Rickford and McNair-Knox (1994) found Audience Design a useful account of their AAVE-speaking informant, where ethnicity and relation to the interviewer significantly favored or hindered the production of certain AAVE features.

2.4.1 Accommodation Theory in SLA

Concerning second language acquisition, Beebe and Zuengler (1983) adopted the framework of Accommodation Theory and extended it to learner speech in order to explain sociolinguistic variation in oral production data obtained from two groups of learners, specifically by manipulating the variable of ethnicity. The first study concerned Puerto Rican children enrolled in either a monolingual or bilingual elementary school in New York City; these children were recorded in interviews in English with an English-speaking Anglo, an English-dominant Hispanic, and a Spanish-dominant Hispanic (I reproduce the labels assigned to each interlocutor's identity by the study authors). There was a correlation in the amount of talk, or *speech quantity* (measured in the number of "T-units" or independent clauses[1]), produced between the children and each of the three interviewers, though this correlation was more significant with the English-speaking Anglo (whom the children may use as a model for their linguistic goals) and with the Spanish-dominant Hispanic (with whom the children may want to assert their distinctive ethnicity). That is, convergence in speech quantity was greatest with interlocutors whom the children viewed as ethnic and linguistic role models. The second study focused on specific linguistic features in the speech of Chinese-Thai children recorded in interviews in Thai with an ethnic Thai interviewer and an ethnic Chinese interviewer. Certain phonological features of Thai contain variants characteristic of (Teochew) Chinese. Despite the fact that both interviewers spoke standard (Bangkok) Thai, the children produced standard Thai features in the interview with an ethnic Thai interviewer but more Chinese variants in the interview with an ethnic Chinese interviewer. In both studies, then, there was evidence of linguistic accommodation toward the identity (whether perceived or actual) of the interviewer in order to assert a sense of ethnic solidarity.

[1] See Loban (1976) for discussion of the "Communication Unit" in oral production, derived from Hunt's (1965) "T-units" ("minimally terminal units") of written production.

In adopting Accommodation Theory to explain their findings, Beebe and Zuengler (1983: 211) were also among the first to highlight the importance of the interlocutor in L2 studies as compared with the target culture: "We are not dealing here with the social and psychological distance of an individual (or group) from the target culture, but with the social and psychological distance of an individual from an interlocutor at a particular situation. We are not focusing on the correlation of an indivudual's [*sic*] L2 proficiency to social and psychological distance from the *target culture* but on the dynamics of social and psychological distance between individual and *interlocutor.*" Beebe and Zuengler also motivate the adoption of Accommodation Theory over a competing SLA model, the Acculturation Model (Schumann, 1978), as follows:

> We agree with Schumann that social and psychological distance from the target culture are the critical factors in second language acquisition, but we believe that in order to fully understand this, we must also study the ebb and flow of feelings of distance/proximity in one controlled situation. Thus, we examine variable performance data […]. As we see it, [Accommodation Theory] could be used to extend Schumann's concept of a relatively fixed social and psychological distance/proximity between an individual and a target culture (at one point in time) to include the idea of fluctuating feelings of social and psychological distance/proximity between one individual and another (at one point in time). (p. 211)

In similar studies, Young (1988, 1991) later extended this distance/proximity between speakers from ethnicity to a collection of attributes, including sex, occupation, educational level, place of origin, and age. Supporting data from L1 Chinese speakers of English showed that the overall degree of social convergence between interlocutors was important in explaining variation in the

speakers' L2 production. For sufficiently proficient learners, social convergence also corresponded to linguistic convergence with the interlocutor, supporting Accommodation Theory.

2.4.2 Priming effects on convergence

Accommodation Theory and Audience Design are both models that can explain the phenomena of convergence and divergence, and these models have been extended to interlanguage variation as well (Beebe, 1985; Beebe & Zuengler, 1983). These models account for the sociolinguistic behavior of modifying one's speech depending on the perception that the speaker wishes to impose on her interlocutor or, more broadly, on a specific social situation. Other models have also been proposed that explain how convergence can be influenced by cognitive mechanisms in addition to psychosocial mechanisms.

The Interactive Alignment model (Garrod & Pickering, 2004) explains convergence by the tendency of speakers to automatically align verbal and non-verbal behavior in a given conversation. Participants are primed by the language forms that are used by their interlocutor(s), which then triggers these participants to favor similar forms in their own production. This alignment is unconscious and non-negotiated, and its benefits include reducing cognitive load as well as establishing common ground among speakers.

This priming caused by interlocutor speech may then be reinforced by priming within an individual speaker. Tamminga (2016) identifies this phenomenon as *persistence*—the tendency to repeat a recently used variant in speech (that is, produced by the same speaker). While the effect on the variables in her study (the *-ing/-in'* alternation and /t/- or /d/-deletion in word-final consonant clusters in English) appears to influence repetition only under specific circumstances (e.g., when the prior instance and the current instance are in the same morphological category and are the same lexical item), the study shows that effects on variation can be produced at the

phonology-morphology interface, and future behavior in an individual speaker may be probabilistically predicted based on the speaker's previously used variants in discourse.

2.5 Interlocutor language background influencing native speakers

Crucially, one interlocutor individual difference not examined in the research outlined in the previous section is the possible effect of the native language background of the speaker's interlocutor; that is, whether the learner is communicating with a native or non-native speaker of the language being spoken. This section will now examine the attention given to this particular social factor in previous literature, first in studies on native speakers; section 2.6 will then follow with studies on L2 speakers.

One way in which the native language of the interlocutor directly influences the speech of certain speakers is in *foreigner talk* or *baby talk* (Ellis, 1985; Ferguson, 1971), which describes the ways in which native speakers (NS) modify their speech in order to be understood by non-native speakers (NNS). Encompassing several linguistic domains, the features of foreigner talk may include a slower speech rate, shorter and simpler sentences, and more attention to pronunciation articulation. This type of language can be examined under Audience Design as an extreme form of style-shifting (in some cases, to such a simplified form of one's native language so as to be rendered ungrammatical), though it may be more accurately described as a special kind of register outside the formal-informal style continuum and devoid of any stylistic connotations. While a speaker using foreigner talk may believe herself to be accommodating to a NNS, the features of this speech do not necessarily take into account the features of one's interlocutor, but rather, perceived elements or *stereotypes* of a NNS's speech (as opposed to *markers* or *indicators*; cf. Labov, 1972).

Style-shifting may be involved in other situations, however, where a divergence strategy involving aspects of Accommodation Theory and Audience Design may be observed in NS-NNS interactions. A NS wishing to maintain greater social distance from a NNS interlocutor may modify her speech to a more normative form of her language if that will distinguish her from a NNS (Zuengler, 1991). Unlike foreigner talk, this form of style-shifting serves to maintain distinctiveness rather than ensure communication with the interlocutor and involves a formal style of the NS's language containing "maximally redundant forms and features" (Valdman, 1981: 44). Perdue (1984) suggests that this normative form of the language may be a conscious attempt by the NS to maintain or reaffirm her identity as a member of a particular speech community and/or as a member of her own ethnic group or culture. Whatever their motivations, then, native speakers can also manipulate sociolinguistically determined features of their speech as a response to the status of their interlocutor as a non-native speaker.

If the status of the interlocutor as a native or non-native speaker influences the way native speakers communicate, one may pose the question whether interlocutor L1 also influences the way NNSs communicate with other speakers, and whether these NNSs are aware of any differences, linguistic or sociolinguistic, in their speech as a result. Second language acquisition research from this angle, however, has been largely limited to the development of the L2 speaker in the classroom setting. The following section will examine such literature that has addressed interlocutor language background as an individual difference, with a view toward its application to non-pedagogical settings.

2.6 Interlocutor language background and SLA

A meta-analysis by Gurzynski-Weiss and Plonsky (2017) has identified several dozen studies in the SLA literature with at least a partial focus on the interlocutor L1 status as influencing

learner behavior. The vast majority of these studies focus on classroom learners and their interaction with their teacher and/or with their peers. Due to the nature of most language learning classrooms, the majority of speaking interactions occur with other learners, and this can be beneficial in facilitating development of the interlanguage grammar, through co-construction and, less commonly, feedback from the other learner (Adams, Nuevo, & Egi, 2011). Inside the classroom, a few studies have examined the status of the second language teacher as NS or NNS (e.g., Gurzynski-Weiss, 2010). There is evidence that NS and NNS teachers promote different outcomes in learners: other teacher characteristics being equal, fluency is increased with NS teachers, while accuracy is increased with NNS teachers (Arva & Medgyes, 2000; Lee, Joo, Moon, & Hong, 2006). Learners also have different beliefs about the teaching abilities of NS versus NNS teachers, which may influence learners' classroom behavior, though other teacher characteristics also play a part in shaping this behavior, and, regardless of actual knowledge or teaching ability, overall the two kinds of teachers have a comparable number of pros and cons according to learners (Benke & Medgyes, 2005; Gurzynski-Weiss, 2010; Jandrey Hertel & Sunderman, 2009; Lasagabaster & Sierra, 2005). Regardless of teacher L1, the nature of learner input from the teacher received a great deal of attention in early SLA studies (e.g., Arthur, Weiner, Culver, Lee, & Thomas, 1980; Long, 1981; Scarcella & Higa, 1982), which found that *teacher talk*, like *foreigner talk*, is characterized by a number of modifications in lexicon, syntax, phonology, and accompanying nonverbal behavior. While syntactic simplifications in teacher talk may overlap with those found in foreigner talk, teacher talk is also likely to include a higher frequency of questions and a variety of interaction devices to maintain the conversation (Porter, 1986) and

prioritize learner comprehension for pedagogical purposes, whereas foreigner talk is likely to prioritize learner comprehension for the specific interactional event between two speakers.

Studies of heritage language learners in academic settings (Potowski, 2004) have shown that restrictions on the choice of the language itself can have psychological implications for learners. Carranza (1995) found that students in bilingual schools experienced a feeling of "pretense" when communicating with each other in one language, knowing that both can communicate more effectively in another language, even if the latter language is not spoken natively (recall, in Chapter 1, a similar interaction between my OPI interviewer and me). Learners may choose to avoid such "pretense" by switching to another language, a divergence strategy observed in both classroom (e.g., Bowles, Adams, & Toth, 2014) and non-classroom (e.g., Giles & Smith, 1979) environments with multilingual speakers. This psychological phenomenon raises the important issue of identity (cf. Dörnyei, 2005) shaping the use of language as defining in-group versus out-group characteristics.

Even when learners maintain use of the L2 under examination, studies examining interactions with other speakers (learners as well as NSs) in experimental settings have found interlocutor influences on the nature of the L2 speech produced by such learners. Many studies that explicitly compare NS-NNS and NNS-NNS dyads[2] (e.g., Varonis and Gass, 1985) have been concerned with negotiation for meaning, finding that NNS-NNS dyads produce more negotiation for meaning contexts than NS-NNS dyads. This is also more typical for mixed-proficiency pairs (e.g., high and low) than same-proficiency pairs (Oliver, 2002); additionally, Oliver found that interlocutor L1 and interlocutor proficiency (but not age or gender) significantly influence the number of negotiations for meaning strategies in learners. Studies have shown that NNSs provide

[2] In this context, a *dyad* is an interaction (e.g., conversation or discussion) between two speakers. Though another person may be present to observe the interaction, only two individuals are speaking.

more opportunities for modified output, but NSs tend to provide more feedback than NNSs, even though NSs are expected to be more tolerant of errors than NNSs (Gurzynski-Weiss, 2010). The nature of this feedback also varies according to interlocutor L1. Bowles et al. (2014) found that language-related episodes (LRE) in L2-L2 interactions overwhelmingly focused on lexical rather than grammatical or phonological difficulties, even though learner self-repair was more common than LREs in both L1-L2 and L2-L2 dyads; on the contrary, in collaborative dialogue, Fernández Dobao (2018) found that lexical LREs tended to be more frequent and more likely to be successfully resolved in L1-L2 than in L2-L2 interaction, even though NSs cannot always be expected to provide the level of assistance that the learner needs for lexical learning. Such findings have led to the postulation that for classroom learners, L2-L2 interactions are qualitatively different from L1-L2 interactions (Mackey, Oliver & Leeman, 2003).

The study by Porter (1983, 1986) specifically analyzed learner speech when manipulating the language background of the interlocutor, and it is likely to be the most similar to the current study in terms of measuring potential interlocutor effects based on the interlocutor's language background. One of Porter's research questions concerned how learner speech changes depending on the interlocutor's proficiency level. She arranged dyadic conversation tasks in English for three groups of speakers: one group of native speakers of English and two groups of L1-Spanish L2-English learners, at intermediate and advanced proficiency levels (determined by TOEFL scores). Porter found that learners talked significantly more to other learners than to native speakers, as measured by total words, suggesting that if production practice is essential for acquisition, learners will benefit from time spent interacting with other learners rather than with native speakers. Importantly, she also found no significant differences in linguistic competence and fluency (in terms of errors and false starts) when learners (at both proficiency levels) talked to advanced versus

intermediate learners. In other words, learners perform the same in terms of grammatical accuracy and fluency whether they are speaking to a more advanced or less advanced learner. On the contrary, however, when expressing opinions, agreement, and disagreement (demonstrating sociolinguistic sensitivity to target-language politeness strategies), learners did not use these strategies to the same extent as native speakers, rendering their speech more "direct" than native speakers; more specifically, learners used many, but not all of the politeness strategies that native speakers used, and at lower frequencies. Note that while the linguistic analysis of grammatical fluency was quantitative, Porter's analysis of learner sociolinguistic competence was qualitative in terms of measuring the use of politeness strategies. Broadly, the finding was that "learners did not provide socioculturally accurate models for expressing opinions, agreements, and especially disagreements" (Porter, 1986: 220). However, in addition to this analysis of politeness strategies being qualitative in nature, it is not clear that Porter compared the use of learner politeness strategies across interlocutor language backgrounds (the question being, simply, "Is the language produced by learners […] appropriate for use in settings outside the language classroom?" (p. 215)). Furthermore, oral production data was based on problem-solving tasks, rather than unprompted conversation, thus limiting the possible scope of topics, and the nature of the tasks focused more on "serious" topics (such as ranking the characters in a story according to their moral behavior).

What we can glean, in any case, from Porter's study and other studies in this section is that learners do not always interact in the same way with NSs compared with NNSs; in other words, the interlocutor L1 variable does play a role in shaping learners' attitudes as well as performance in second language learning. It is thus reasonable to assume that as they exit the classroom environment and enter the broader target language community (that is, for those learners

continuing to use their L2), the speech of these learners may continue to be influenced by the L1 status of the interlocutor.

As one might expect, the above studies have been largely concerned with linguistic, rather than sociolinguistic, development in learners. While Porter's study looked at linguistic *and* sociolinguistic performance, the qualitative sociolinguistic analysis did not focus on specific sociolinguistic variables, or those that vary according to style (with the exception of one category of informal speech when expressing opinions). Few studies, for example, examine feedback (such as recasts) on sociolinguistically conditioned variation in classroom learners, due in large part to the lack of opportunities to interact with target language speakers in different sociolinguistic contexts. In the classroom, explicit instruction may have focused on the communicative contexts in which to use certain forms that can be opaque to beginning learners (such as 2SG address pronouns *tu* versus *vous* in French), but most of the time, primacy is given to correct grammatical expression (focus on form) or successful communication by the learner (focus on meaning), regardless of the sociolinguistic context. In addition, classroom textbooks often gloss over sociolinguistic variation or even promote infelicitous structures (Etienne & Sax, 2009). Moreover, classroom-based learning largely favors the use of formal (written) variants.[3] Thus, it is also reasonable to assume that while learners will enter the target language community with some metalinguistic knowledge of certain sociolinguistic variables, they will tend to overuse formal variants in initial interactions with native speakers, given that learners typically have little (or no) practice identifying the contexts in which one might find variation, and just as importantly, they may have not yet had any feedback from speakers (native or otherwise) on which variant is appropriate in each communicative context.

[3] Despite efforts such as those by Valdman (1989) to shift the L2 acquisitional focus from L1 idealized forms to more naturalistic norms; see also Kramsch (2002) and Orton (2014).

In sum, while classroom-based research is certainly useful in measuring the impact of instruction, the rather specialized communicative environment of the SLA classroom does not easily lend itself to development of the full range of sociolinguistic abilities in the L2 learner; moreover, many classroom learners lack the grammatical proficiency for sustained interaction in the target language. It is unsurprising that sociolinguistic ability is generally acquired in the later stages of SLA, after a certain level of grammatical competence has been acquired (Adamson, 1988; Young, 1991). The review of SLA studies in this section concerning interlocutor language background reveals a dearth of literature focusing on this sociolinguistic acquisition. Thus, we may be able to broaden our understanding of this acquisition process in the context of interlocutor individual differences if we look beyond the classroom at learners living in the target language community, in conversations between speakers who are capable conversation partners, while extending the inquiry of interlocutor effects on sociolinguistic performance to a range of sociolinguistic variables.

2.7 Sociolinguistic variables and study abroad

Study-abroad learners are one category of speakers that can fulfill the requirements outlined above. Many studies have shown that a stay in the target language community (and especially, contact with native speakers) leads to an increase in proficiency in all areas of linguistic ability, including acquisition of sociolinguistic competence (recent examples for French include Kennedy Terry, 2017; Regan, Howard, & Lemée, 2009; Sax, 2003; examples for Spanish include Geeslin, Fafulus, & Kanwit, 2013; Geeslin, García-Amaya, Hasler-Barker, Henriksen, & Killam, 2012; Knouse, 2013). Importantly, benefits of study abroad also include sociolinguistic and sociocultural knowledge (see Freed, 1995, for examples therein), as well as recognition of sociolinguistically appropriate variation. Study-abroad learners also become aware of regionally

marked variation, as evidenced by a large research program for L2 Spanish headed by Geeslin (Geeslin et al., 2013; Geeslin, Garcia-Amaya, Hasler, Henrikson, & Killam, 2010; George, 2013, 2014, 2018; Kanwit, Elias, & Clay, 2018; Kanwit, Geeslin, & Fafulas, 2015; Kanwit & Solon, 2013; Knouse, 2012; Linford, 2016; Ringer-Hilfinger, 2012; Salgado-Robles, 2011), though learners' actual production tends to remain closer to non-regional, classroom-based norms (and despite some researchers advocating for instruction of regional variation, learner usage of certain forms may not be received favorably in all NS communities; cf. Auger & Valdman, 1999). In any case, Regan et al. (2009: 134) summarize this benefit of study abroad as follows: "[A]fter a year abroad, the L2 speakers approximate L1 variation speech patterns. This approximation is closer in relation to some variables than others, but in general, the speakers are using variation patterns which are significantly more similar to those of native speakers than before they went abroad and more than those of speakers who do not go abroad." Longer stays in the target language community can also yield L2 proficiency gains that render certain learners' linguistic abilities essentially nativelike (Birdsong, 1992; Donaldson, 2008; Forsberg, Bartning, Engel, Gudmundson, Hancock, & Lindqvist, 2014), though measures of sociolinguistic ability have received comparatively less attention in this L2 population, further motivating the inclusion of an additional group of highly proficient, near-native learners to compare with a group of study-abroad learners.

2.8 Type I and Type II variation

When measuring sociolinguistic ability in learners, it is also important to make the distinction between what has been referred to in the literature as Type I and Type II variation (cf. Adamson & Regan, 1991; Bayley & Regan, 2004; Mougeon, Nadasdi, & Rehner, 2010; Rehner, 2002, 2005). Type I variation (also called *vertical* variation) refers to variation that exists only in the L2 language or the *interlanguage*. Variants of this kind would include (morpho)syntactic forms

that deviate from the grammar(s) of NSs of the language. Thus, learners may alternate between targetlike forms (the NS model) and one or more non-targetlike forms that are not attested in NS grammars (apart from performance mistakes, such as slips of the tongue). Non-targetlike forms produced by the learner may be influenced by a variety of acquisitional factors such as transfer from the learner's L1, incorrect application or overgeneralization of a learned (morpho)syntactic rule, or processing or working memory constraints.

Type II variation (also called *horizontal* variation) concerns alternation of forms that exist in the speech of NSs of the target language. In these cases, a certain variant may be favored by NSs more frequently due to various social, linguistic, and contextual factors, and use of an alternative variant may be considered stylistically infelicitous, but all variants conform to the grammar(s) of the NSs (or at least a subset thereof). Conversely, the notion that all variants are grammatical may not be shared by all speakers, due in part to prescriptive pressure in many language communities to conform to a standard language form, which is usually based on the written form and carries prestige within this community (and where infelicitous use of a particular variant may be considered by certain NSs as "ungrammatical" even if the variant is considered appropriate in other contexts).

Type II variation can be divided into two subsequent categories: free variation and rule-governed variation. Free variation implies a choice between multiple variants that does not correlate with any linguistic, social, or psycholinguistic factors. Matthews (1997) gives the example of flapped and continuant [r] in some speakers of English. However, since there may potentially be an infinite number of variables influencing the choice of a variant, "true" free variation may only be a temporary way of explaining the lack of correlation to specific factors, until such factors are identified. Ellis (1999) also explains that free variation plays a different role

in the L2 acquisition process, where a learner, upon acquiring variation in the input—but not the subtle social meanings distinguishing the use of multiple forms—may freely alternate among these forms until the learner's social situation changes and causes her to initiate reflections on the contextually appropriate use of these forms; if the learner fails to resolve these differences, fossilization may occur. Rule-governed variation, on the other hand, involves the probabilistic selection of a variant influenced by a combination of linguistic and extralinguistic factors. The challenge for the language learner, of course, is identifying the relative importance of each of these factors, and for which both the linguistic and social factors are likely to differ from the factors conditioning variation in the learner's L1. As Rehner (2005: 15) puts it, success in the use of Type II variation involves: "(a) learners' use of the same expressions as NS; (b) their use of such expressions at levels of discursive frequency similar to those found in the speech of NS in the same situation; and (c) the correlation of such uses with similar independent factors, both social (e.g., social class, sex, and style), and linguistic (e.g., the surrounding lexical and syntactic context), affecting the uses by NSs." As with Type I variation, transfer or overapplication of L1 parameters, or working memory constraints, may also influence Type II variation, even if such influences have been less commonly studied (but see, e.g., Zahler (2018) for working memory constraints on L2 sociolinguistic variation). Thus, caution must be observed when comparing the factors influencing L1 and L2 sociolinguistic variation; superficially similar patterns of variation may be the result of different underlying mechanisms between L1 and L2 grammars.

With these considerations in mind, Type II variation involving rule-governed selection of forms is the focus of interest in the current study. The types of learners to be examined would include speakers at a proficiency level where Type I variation is not expected to prohibit spontaneous and sustained comprehensible oral production, but where these speakers have

encountered variation in their input—whether or not they have initiated reflections on the social and stylistic value of Type II variants, and whether or not they have begun to use these variants in the same distribution patterns as members of the target-language community.

As introduced in section 2.7, studies on study-abroad students have shown that production of sociolinguistically conditioned variables generally becomes more nativelike as proficiency increases and as learners become more aware of social contexts in which to produce the appropriate variants (e.g., Regan et al., 2009). However, deviation from nativelike sociolinguistic norms (as measured in the use, and frequency of use, of certain sociolinguistic forms) can persist even in learners demonstrating advanced grammatical and/or phonological proficiency. Why might this be the case?

Some researchers question the futility of learners aspiring to a native-speaker ideal. For example, Cook (2005) says that learners are misguided if their goal is to speak as much like native speakers as possible. What is important is that the goals of the language learner must be specialized for each learner. Thus any study of L2 proficiency should include a component (such as in a language background questionnaire) having the learners specify their goals for study in the L2. Classroom learners may simply be studying the L2 in order to meet certain academic requirements, while other learners may opt for a more passive knowledge of the language—for example, sufficient comprehension to read in the L2 or watch/listen to media in the L2. Some cases of targetlike deviation may be explained by these factors. Lennon (1993) emphasizes that learners will vary individually concerning their own ideal terminal state of proficiency, reflecting variation in importance attached to different communicative goals, such as, according to Lennon, "the weight attached to accuracy compared to fluency, to getting one's message across compared to sounding nativelike" (p. 41). Different learners will also assign more weight to one domain of

language over another, prioritizing phonology, vocabulary, or syntax in different combinations. Citing Kohn (1982), Lennon further argues that "any assessment of the advanced learner should take into account not only current proficiency and native speaker norms, but the learner's own goals" (p. 41).

Regardless of learners' specific motivations and the relative importance they assign to specific domains of language production, one can presume that most learners wishing to maintain sustained interaction with other speakers of the target language would want to, minimally, achieve L2 "fluency"—though what this entails may be defined differently by each learner (with the notion of "fluency" almost certainly defined differently by learners compared to definitions in SLA literature (cf., e.g., Forsberg Lundell et al., 2014)). For most learners sojourning in the target language community, however, this likely means achieving a level of fluency allowing them to carry out communicative functions in a variety of social situations, which will likely differ in terms of formality and the relationship between the learner and her interlocutor (whether the interlocutor is a classmate, a bank teller, or a friend's child, for example; cf. style-shifting in Bell's (1984) Audience Design framework). Segalowitz (1976) hypothesized that learners will experience communicative difficulties "when the socio-linguistic demands of the situation require them to use a speech style outside their repertoire" (p. 129); moreover, Segalowitz found that native French adolescents had a less positive perception of L2 French adolescents when these learners were required to speak in a more casual manner compared to a formal manner. As mentioned in section 2.1, the ability to "live" in a second language requires a certain level of sociolinguistic proficiency (Geeslin & Long, 2014); as Segalowitz (1976: 130) notes, each communicative interaction with native speakers "carries its own sociolinguistic demands and, until the speaker learns how to handle them, he may shy away from those that make him uncomfortable." Van Compernolle and

Williams (2012a: 237) characterize the ability to handle these demands as *sociolinguistic agency*, which redirects the focus from a native speaker ideal to the learner's ability to recognize how the use of a particular variant "simultaneously reflects and creates the context in which it is used, is a performance of one's social identity at the time of utterance, and affects one's environment and interlocutor(s)." We can reasonably assume, then, that learners living in the target language community want to acquire sociolinguistic competence even if they are not explicitly aware of how this is accomplished or whether it can be separated (operationally or theoretically) from achieving L2 fluency.

From a psycholinguistic perspective, studies have shown that, even if learners have metalinguistic knowledge of stylistically appropriate forms, more limited L2 processing capabilities may prevent them from producing appropriate forms when the production task is cognitively demanding (cf. Housen, Kuiken, & Vedder, 2012; Lyster, 1993). Thus, the question arises as to whether learners make conscious choices from among competing stylistic variables. Lyster (1993) found that L2 French immersion learners overused informal *tu* in spontaneous production tasks when formal *vous* was expected, suggesting that the message (e.g., focus on meaning) took primacy over stylistic appropriacy; the cognitive demands of selecting the appropriate form in an online task may lead learners to demonstrate less nativelike sociolinguistic performance than their metalinguistic knowledge would otherwise indicate. French and Beaulieu (2016) also found that unplanned discourse in L2 French learners led to more nativelike use of one sociolinguistic variable (*ne*-deletion) but not another (/l/-deletion), which may be explained by the relatively more complex linguistic constraints governing the latter variable, requiring additional cognitive effort. Furthermore, French and Beaulieu also provide evidence that the selection of overly formal variants was a conscious choice by certain advanced learners; they did not always

deem it necessary to try to speak exactly like native speakers, given their status as learners, and they felt that they could be better understood by native speakers when adhering to standard forms. Thus they adopted a style that they deemed better reflected their identity as a language user, even though it was not necessarily the same style adopted by their interlocutors.

The question concerns, then, on what level learners remain aware of their stylistic choices in a communicative context, and how these choices may be perceived by their interlocutor(s). *Awareness* has been operationally defined in multiple ways in SLA literature (cf. Leow, 2001, for an overview), drawing upon the "noticing hypothesis" (cf. Schmidt, 1990), where individual learner differences in awareness of a target language form (such as the experience of some cognitive content or external stimulus) may lead to changes in learners' behavioral patterns and subsequent developmental paths in L2 acquisition. As may be expected, much of this research has focused on external awareness of grammatical forms such as morphological rules. Less attention has been given to learners' awareness of their own stylistic choices, such as in French and Beaulieu (2016), which involved written questionnaires concerning learner reflections on specific stylistic variables. Analysis of these questionnaires also indicates that, within the span of a single communicative interaction, even learners at advanced levels may not adapt their use of these variables to the level of formality appropriate to the interaction, even when their interlocutor may use a range of informal variants. This question also concerns whether speakers make a conscious choice for each variant produced. In learners, Howard (2012) has addressed this question as follows: "…to what extent are [learners] genuinely consciously aware of the informal stylistic value that the [sociolinguistic] variables convey, and in what way do they purposely adopt such a style as a means of sounding increasingly native-like? Or, are they simply using such informal variants as formulaic routines without being fully aware of their inappropriateness in some stylistic

contexts?" (p. 31). The use of formulaic routines implies less conscious selection of stylistic forms—again, possibly due to cognitive demands from competing linguistic domains. Furthermore, the intentional adoption of a style in order to sound increasingly nativelike falls within the scope of interlocutor effects; that is, with which types of interlocutors would learners wish to sound increasingly nativelike? Are there learner individual differences regarding the types of interlocutors for whom the demonstration of more nativelike speech is desired?

With such observations in mind, let us turn to how sociolinguistic ability in L2 speakers can be measured. Specific variables must be chosen for study that are conditioned by the social situation and that are measurable in the language input as well as in the L2 speaker's production. Which sociolinguistic variables fit these criteria, and in which language(s) are these variables particularly salient? The following section will treat these questions.

2.9 Sociolinguistic variation in L2 French

Researchers (e.g., Gurzynski-Weiss, 2010) have noted that what is known about interlocutor individual differences in SLA largely comes from studies of L2 English, and that it is important to expand the discussion to other languages to see if these differences also influence speaker behavior in other L2 contexts. Concerning sociolinguistic variation, certain structures have been studied cross-linguistically, such as pronominal subject expression in Chinese (e.g., Li, 2017) and Spanish (e.g., Geeslin & Gudmestad (2011), and forms of address in French (e.g., Dewaele, 2004b; van Compernolle, 2015) and Spanish (e.g., van Compernolle, Weber, & Gomez-Laich, 2016), though the sociolinguistic variables examined are likely to be subject to different linguistic and social factors for each language.

As previously cited studies have indicated, French contains sociolinguistic variables in which the level of formality of one's speech can be reflected in the selection of a particular variant,

such as the inclusion of the negation particle *ne*, or interrogative subject-verb inversion, in formal contexts. Concerning acquisition of sociolinguistic competence, it is pertinent that the codification of Modern French has resulted in particularly strong sociolinguistic constraints on these variables (Lodge, 2007). With written French maintaining high prestige for historical and cultural reasons, formal features largely align with the written form of the language. Spoken French, on the other hand, is characterized by a number of informal variants, which may also be stratified to multiple levels of formality within the spoken language itself, and for which appropriate knowledge of style-shifting is necessary.[4] Due in part to the high prestige of the written (formal) form, Dewaele (2007: 20) states, "Acquiring sociolinguistic competence in French equates the ability to navigate a social minefield," but, "What is a minefield for learners is a rich area of investigation for researchers." Therefore, in this section, I outline the influential developments in sociolinguistic variation research on L2 French. Note that this discussion constitutes a broad overview; some of the studies mentioned here will be revisited in more detail in Chapters 4 and 5 as they relate to the variables to be analyzed in the current study.

Much of the variationist research in L2 French has been carried out in Francophone communities in Canada. For example, Mougeon and his colleagues have examined L2 sociolinguistic variation in French immersion contexts in Montréal, recording interviews with Anglophone high school students enrolled in these immersion schools (Mougeon, Nadasdi, & Rehner, 2002, 2010; Mougeon & Rehner, 2017; Mougeon, Rehner, & Nadasdi, 2004; Nadasdi, Mougeon, & Rehner, 2003; Rehner & Mougeon, 1999, 2003). The sociolinguistic variables analyzed in this extended research network include studies on *ne*-deletion, interrogative structures,

[4] There is debate over the question of French speakers having one grammar with stylistically influenced variation or two grammars in a situation of diglossia (involving distinct formal and informal forms of the same language); see, e.g., Massot (2010), Rowlett (2013), and Zribi-Hertz (2011).

on versus *nous*, phonological elision (schwa deletion and /l/-deletion), future forms (e.g., *je vais partir* 'I'm going to leave' versus *je partirai* 'I will leave'), discourse markers (such as *alors* and *donc*), and informal lexicon. Questions posed by this research group have included whether immersion students use the same range of variants (and at the same frequency) as native speakers, whether these students are conditioned by the same linguistic and extralinguistic factors on L1 variation, and to what extent L2 sociolinguistic variation is influenced by factors inherent to the language learner's profile: her native language, her degree of contact with NSs, and her exposure to variation in classroom settings (such as variation in the language textbooks or variation used by classroom instructors). Regarding the last point, Mougeon and colleagues have found that while instructor speech may be more nativelike than textbook input with regard to sociolinguistic variation, these instructor models still tend to overproduce standard forms compared with the native speech communities at large. Overall, learners still tend to maintain a smaller sociolinguistic "repertoire" than NSs and an overreliance on marked standard forms. However, frequent interaction with members of the target language community has been shown to have a positive correlation with increasingly targetlike use of informal variants. Moreover, there is some evidence for certain extralinguistic factors conditioning the same sociolinguistic variables in learners as in NSs; for example, female learners and middle-class learners demonstrated higher frequency of use for the more formal variants compared with male and working-class learners, a tendency found in L1 French speakers (cf. Blondeau & Nagy, 1998, for subject doubling; Rehner et al., 2003, for *on/nous*).

Dewaele, Regan, and colleagues (Dewaele, 1999, 2002, 2004a, 2004b, 2005, 2007; Dewaele & Regan, 2001, 2002; Regan, 1995, 1996; Regan, Howard, & Lemée, 2006, 2009) have also conducted many studies on L2 acquisition of European French, primarily with learners in

study-abroad contexts, and focusing on a similar range of sociolinguistic variables, such as *ne*-deletion, *on* versus *nous*, and interrogative structures, in addition to pronouns of address. These studies also point to authentic interaction with NSs as crucial to developing style-shifting competence leading to more nativelike variation. Other forms of authentic NS input, such as contact with French-language media, were also linked to more targetlike variation in some cases (e.g., Dewaele & Regan, 2002, for *ne*-deletion). There is also evidence of a "pendulum effect," in which learners initially overuse the formal variant, followed by over-use of the informal variant even in formal contexts, in an attempt to sound "native" (Dewaele, 2002). In other cases, following a period of study abroad, usage of these variables has been shown to begin reverting back to pre-study-abroad tendencies, highlighting a potential need for reinforcement from authentic NS input. Finally, Dewaele (2004) is among the first to report on evidence of an interlocutor effect of language background on *ne*-deletion, with advanced learners demonstrating higher *ne*-deletion when speaking with other learners than when speaking with NSs.

Sax (2003) focused on groups of learners at three L2 proficiency levels and found that length of study abroad was also a significant factor in developing stylistic variation. Learners in the long-term study-abroad group (from more than several months to several years) demonstrated more nativelike variation than the short-term study-abroad group (about two to five months), and both groups outperformed the learners with little to no study-abroad experience (0-2 weeks) over a range of variables (*ne*-deletion, interrogative structures, *on/nous*, /l/-deletion). Data from the lower-proficiency group point to *on/nous* as the first structure with evidence of variation in learners. Most highly proficient learners show evidence of variation for all variables, though this group still displayed lower *ne*-deletion and less stylistic variation overall compared with a NS control group.

Howard (2012) also examined multiple variables appearing frequently in the production data of L2 learners: *ne*-deletion, /l/-deletion, *on/nous* usage, liaison realization, and expression of futurity. Howard sought to determine how these variables may be linked to each other—that is, on an individual level, whether use of a specific sociolinguistic variable necessarily entails the use of another variable, and which variables are more easily acquired in learners. The learners in this study were at a higher proficiency level (identified as "advanced") than in many earlier studies, and they approached nativelike variation in analysis of some variables (such as *ne*-deletion and *on/nous*)—even for some learners who had not had extended exposure to native speakers. However, with other variables, such as futurity and /l/-deletion, their production was much less nativelike. This inconsistency across a range of variables led Howard to question why some sociolinguistic variables are easier to acquire than others, whether learners are aware of this imbalance, and how this imbalance might sound to NSs.

Extending the inquiry to highly advanced learners, Donaldson (2008) examined dislocation structures in near-native speakers of French. He found evidence that certain highly advanced speakers are capable of coordinating syntactic and pragmatic knowledge, and that their mastery of left- and right-dislocation was on par with native speakers. Using data from the same speaker group, Donaldson (2017) also conducted a variationist analysis of *ne*-deletion and found that near-natives as a group were not statistically different from native speakers; however, about half of the near-natives overused *ne* compared with natives, and Donaldson notes the influence of "nonnative pragmatic conservatism" that may still persist at near-native proficiency levels with regard to sociolinguistic behavior. Nevertheless, his evidence points to the "possibility of nativelike acquisition of a sociolinguistic variable governed by a complex interplay of factors by at least some near-native speakers," where such learners had acquired more than a superficial

understanding of *ne* usage (deleting *ne* in speaking but retaining it in writing) with evidence of sensitivity to the same linguistic and sociostylistic factors underlying NS usage patterns of this variable (p. 165).

A research program by van Compernolle and Williams (van Compernolle, 2008, 2015; van Compernolle & Williams, 2009, 2012) has also examined L2 sociolinguistic development through various forms of electronic communication by learners, such as chat and text messages, as well as distance-based methods for examining L2 sociopragmatics (such as web-based surveys eliciting responses for appropriate use of pronouns of address). Van Compernolle and Williams (2012b) found a longitudinal increase in informal *ne*-deletion in intermediate learners' chat-based communication but less clear evidence of a benefit of metalinguistic instruction regarding sociopragmatic uses of *ne* (e.g., emphatic use of *ne* in informal contexts); they thus argue for the introduction of classroom-based variation at the earliest stages of L2 education. The authors also report on an anecdotal interlocutor language background effect; though they could not control for the native language of the participants who chatted with their learners, they observed that some learners only began to attempt variation when a native French speaker entered the chat.

In an examination of L2 sociopragmatic development, van Compernolle (2015) found that, in an offline task choosing between *tu*/*vous* for forms of address, advanced and intermediate L2 speakers (though not necessarily near-natives) were more conservative, and less committed to their *tu/vous* choices, than native speakers. That is, learners erred on the side of caution by being more polite (using *vous* when natives tended to use *tu*), due to learners' status as "cautious outsiders" relying more heavily on learned rules related to formality, distance, and power. However, in online speaking tasks, as noted previously in Lyster (1993), learners may overuse informal variants. Etienne and Sax (2009: 588) also point out that "systematic use of informal variants by nonnative

speakers can be judged negatively by native speakers, so that students may not integrate or may even be ridiculed" (see also Ryan & Giles, 1982; Swacker, 1978; Valdman, 1992). Moreover, highly proficient non-native speakers may wish to avoid informal variants, consequently creating "a false impression of incomplete competence" (Dewaele, 2007: 4), which may be linked to Donaldson's (2017) observation of non-native pragmatic conservatism in his near-native speakers.

This overview of L2 sociolinguistic variation in French is not meant to be exhaustive; however, certain general observations can be made. To summarize, studies on sociolinguistic variation in L2 French have identified a strong tendency for learners to overuse formal variants where NSs tend to use informal variants. Naturalistic exposure (authentic interaction with NSs and, to a lesser extent, exposure to French-language media) reappears as an influential factor in more targetlike usage of these variables, even if few learners, even highly proficient ones, display nativelike patterns. Furthermore, while studies on study-abroad learners are abundant, there is, aside from Donaldson (2008), a dearth of large-scale studies on the acquisition of sociolinguistic variation in highly proficient, near-native speakers of French.

2.9.1 Sociolinguistic variation in L2 French: The question of interlocutor L1 effects

As we have seen in the previous section, studies of sociolinguistic variation in L2 French have included a wide range of learners both in and outside the classroom. As may be expected based on Gurzynski-Weiss and Plonsky's (2017) meta-analysis, however, interlocutor language background has received little attention in previous studies. A non-exhaustive sample of studies on variation in L2 French (see Table 2-2), including many studies referenced in the preceding section, reveals differences in the kinds of interlocutors selected for eliciting speech from L2 French learners: whether due to convenience or other factors, native or non-native speakers may

be involved (sometimes with the researcher playing the role of interlocutor), but often without explicit justification for this methodological decision.

Table 2-2. Interlocutor L1 in studies of sociolinguistic variation in L2 French

Study	Sociolinguistic variables	Interlocutor L1
Painchaud, d'Anglejean, & Vincent (1982)	*ne*-retention[5]	NS
Trévise & Noyau (1984)	*ne*-retention	NS
Regan (1996)	*ne*-retention	NNS (author)
Thibault & Sankoff (1997)	*ne*-retention	NS
Rehner & Mougeon (1999); Rehner, Mougeon, & Nadasdi (2003)	*ne*-retention; *on/nous*	NS
Dewaele (1999)	interrogatives	NNS and NS
Dewaele (2002)	*on/nous*	NS (author)
Dewaele & Regan (2002)	*ne*-retention	NS (author #1; bilingual French-Dutch)
Nagy, Blondeau, & Auger (2003)	subject doubling	NS
Sax (2003)	/l/-deletion; *ne*-retention; interrogatives; *on/nous*	NS
Dewaele (2004a)	*ne*-retention	NNS and NS
Thomas (2004)	*ne*-retention	(no interlocutor)[6]
Tyne (2004)	informal vocabulary	NNS
Uritescu, Mougeon, Nadasdi, & Rehner (2004)	schwa deletion	NS
van Compernolle & Williams (2009)	interrogatives; *on/nous*	NNS[7]
van Compernolle & Williams (2012b)	*ne*-retention	NNS (some presumed NS)
Howard (2012)	*ne*-retention; *on/nous*; /l/-deletion; liaison; futurity	NNS
Kennedy (2012)	/l/-deletion	NNS
French & Beaulieu (2016)	*ne*-retention; /l/-deletion	(no interlocutor)[8]
Donaldson (2017)	*ne*-retention	NS

[5] Though the majority of studies on French negation have analyzed the *ne* particle in terms of deletion, I will refer to this variable as *ne-retention* and report all statistics in these terms, as some authors have done (Armstrong, 2002; Ashby, 2001; Donaldson, 2017), unless otherwise indicated. Other studies reporting in terms of *ne*-deletion will have the relevant statistics converted in order to facilitate comparisons across studies.

[6] Learners had no interaction with an interlocutor, as they recorded their oral responses to pre-recorded topic prompts; the author does not specify whether the participants were provided an explanation for a target audience to whom they were to address their recordings.

[7] The two van Compernolle and Williams studies (2009, 2012b) concern text-based online chat discussions. As in Thomas (2004), the lack of face-to-face interaction with an interlocutor renders these studies less relevant for comparison.

[8] Like Thomas (2004), learners recorded their oral responses to pre-recorded topic prompts.

As this table shows, studies vary according to the status of the interlocutor eliciting speech from the learners—the majority use a native speaker, while several others use a non-native speaker. In addition to inconsistent reporting in the description of the interlocutor(s) in these studies, there are varying amounts of detail in the justification for selecting the interlocutor(s) for each study, and there is not necessarily an identifiable trend toward more transparency in more recent studies.

In several cases, there is no mention of the native language of the interlocutor(s). In Painchaud et al. (1982), the interlocutor was a linguistic anthropologist whose L1 is presumed to be French, while in Trévise and Noyau (1984), the only reference is to an "observer" who is presumed to be one of the authors. For Uritescu et al. (2004), no interlocutor information is given, but data is based on a corpus (cf. Mougeon & Beniak, 1991) where the interlocutors were native speakers. Other studies (Dewaele, 1999; Rehner & Mougeon, 1999; Rehner et al., 2003) simply mention the interlocutor's status as native French speaker along with other personal characteristics, including those cases where the interlocutor is one of the researchers, as in Dewaele (2002, 2004a) and Dewaele and Regan (2002). Since differences in formality were a central component of her study, Sax (2003) provides a more thorough description of her two NS interlocutors (even including a picture of them in the appendix section of her dissertation), where the appearance of the interlocutor and the setting of the conversation contributed to the desired perception of a highly formal or informal interaction; it is implied that native speakers were chosen so that their data could be compared to previous L1 studies on Metropolitan French.

When the researcher is a non-native speaker of French who participates as an interlocutor, attention to the interlocutor's characteristics does seem to be a concern. Regan (1996: 185) does not comment on her NNS status but does acknowledge her institutional role as affecting the formality of her interviews: "The speakers, students in the department where I taught, knew me to

be a member of the faculty but were not students in courses I taught. Prior to the study, I was not acquainted with them, but, over the period of the study, a certain friendship was established. In their use of language, then, over this time, it is possible that the increased intimacy led to a slight decrease in formality." It is worth noting that this slight decrease in formality may directly impact the use of variables prone to style-shifting—in other words, precisely the variables in question in her study. Kennedy's dissertation (2012) devotes a paragraph to her interlocutor status as a near-native speaker of French and as a graduate student from an American university (her 2017 article on /l/-deletion data, published under the name Kennedy Terry, does not include these details). Like Regan, Kennedy mentions that it was easy to establish a rapport with her participants, by focusing her sociolinguistic interviews mainly on the study-abroad experience; she does mention that she engaged in some speech accommodation with less proficient learners, which included a lower speech rate and few informal linguistic and sociolinguistic variants. Tyne (2004) devotes some attention to his NNS status in his interactions with his participants; though he never emphasized this NNS status and carried out all interactions with his participants in French, he acknowledges the possible effect of being perceived as a "foreigner" in influencing his participants' behavior. Finally, Howard (2012: 23) makes a brief judgment, albeit indirectly, on his own French-speaking status: "Although not a native speaker, the interviewer demonstrated near-native competence in French." The presupposition is that a native speaker would have been preferred, but one was not recruited for unexplained reasons.

It is evident that for many of these studies, the choice of the interlocutor rests upon the availability of a suitable conversation partner, which is dependent upon time and resource constraints. Apart from Regan (1996), Sax (2003), and Tyne (2004), the choice of interlocutor as having a confounding effect on the variables studied has received little attention. Moreover, several

of these studies involve conversational interaction between a university professor/researcher and a student who is an L2 learner of the language spoken in this interaction. In addition to the inherent power imbalance involved in this type of dyad, one can surmise that differences in age and social class (or profession) are also present in many cases, even when an interlocutor not involved in the research aspect of the study has been recruited (this type of participant is often referenced in the SLA literature as a *confederate*). Thus, it seems that if the social divergence between the interlocutor and the study participant has an effect on the performance of the participant, this divergence is often accepted simply as an artefact of the study design. However, it is unsurprising that, given the increased social divergence, L2 learners overuse the formal or standard variants of the sociolinguistic variables in question, presumably influenced by perceptions of what is "accurate" or "grammatical" and thus producing speech that may not reflect the learner's actual sociolinguistic abilities when compared with native speaker norms.

Some researchers, such as Dewaele (2002), mention organizing the recording environment so as to produce a relaxed environment for the participants. This consideration is certainly an important part of any study design, as people are understandably concerned about performing well even when being observed speaking in their native languages, let alone in a second language, and a non-threatening speaking environment presumably leads to more naturalistic (and likely more nativelike) production. Again, we must keep in mind that no study can completely avoid the observer's paradox unless unethical measures are taken (such as surreptitious recording of conversations, which violates responsible conduct of research). However, any study comparing learner speech with NS norms (and attempting to explain learner deviation from targetlike norms) would want to be concerned with interlocutor characteristics influencing an environment where the most favorable context for L2 production of targetlike forms is sought.

As introduced in the previous section, the only study in the above list that specifically investigates interlocutor L1 as a potential factor (among others) in influencing sociolinguistic variation is Dewaele (2004a).[9] In this study, Dewaele found that learners in interaction with native speakers (NNS-NS) produced lower rates of *ne*-retention (53.5%) than in interaction with other learners (NNS-NNS), 75.5%. Dyads with different interlocutor L1s had significantly less *ne*-retention than same-L1 dyads; dyads with age differences had marginally higher *ne*-retention compared with same-age dyads; and dyads with gender differences were not significantly different from same-gender dyads.

According to Dewaele, then, for L2 learners, interlocutor L1 status may be more important than age or gender differences when it comes to sociolinguistic variation. This is significant, because no other sociolinguistic studies of French have focused on this extralinguistic factor, and there has been scant attention to this factor on studies of other languages, with the exception of Porter (1986), who did not examine specific sociolinguistic variables. Most L2 studies (and nearly all L1 studies) consider the possible effects of age and gender of the *speaker* (especially in *ne*-retention studies), but not of the interlocutor (though see Biers, 2014, for these interlocutor effects in L1 French). Moreover, the interlocutor variable may turn out to be even more significant if the effects of these other social factors can be eliminated or minimized through the design of the study. Considering the inconsistencies in the selection of interlocutors in the L2 French studies described in this section, it is plausible that Dewaele discovered statistical evidence for an overlooked feature (if not a flaw) in the methodological considerations of studies on L2 sociolinguistic variation, even with Dewaele's caveat that his study's unequal sample sizes and multiple statistical analyses of the same data require caution when drawing definitive conclusions.

[9] Dewaele's 1999 study on interrogatives also had NS-NNS and NNS-NNS dyads, but no between-group analyses were reported, nor was interlocutor L1 analyzed as a factor in variation.

Retention of at least the NS versus NNS interlocutor component would thus be in order for any future studies, if we wish to shed more light on the possible effect of interlocutor language background. Two other significant considerations regarding Dewaele's study warrant attention as well.

First, Dewaele obtained one recording for each speaker; there were learners who interacted with L1 French speakers and learners who interacted with other learners, but no speaker who interacted with *both* kinds of interlocutors. Any effect of interlocutor language background would be more robustly measured by comparing the *same* speaker across different interlocutors; that is, by measuring intra-speaker in addition to inter-speaker variation, again keeping in mind that isolation of this variable is better achieved by minimizing other social variables such as age, gender, or social class across one speaker's interlocutors.

Second, Dewaele's justification for differences across interlocutors rests upon a *t*-test measuring between-group equality of means. This statistical analysis would provide an adequate explanation of variation if speakers produced negation contexts in identical ways across interlocutors. However, we know that the spontaneous nature of oral conversation renders exact comparisons essentially impossible. A variationist analysis, which considers the possible effect of multiple linguistic and social factors, would provide more robust evidence of an effect of interlocutor when accounting for the large number of factors, both linguistic and social, that influence the speech choices of native and non-native speakers. As seen earlier in section 2.9, many studies have conducted variationist analyses to determine the significant factors influencing variation for a range of variables. Since such an analysis will be relevant for teasing out the potential relative influence of interlocutor effects, a deeper review of variationist approaches to

the broader field of SLA is outlined in the next section, after which I will return to the question of the interlocutor effect and its implications on the methodology of the current study.

2.10 Variationist approaches to SLA

As introduced in section 2.2, Bayley (2005) points out that early research on variation in SLA (among them, Beebe, 1977; Ellis, 1987; Tarone, 1985) attempted to explain learners' linguistic choices by reference to a single co-occurring contextual factor. It is perhaps not coincidental that in each of these studies, the factor that explained the interlanguage variation supported the authors' theoretical positions in each case (such as speech accommodation, planned versus unplanned discourse, discourse domain, and attention to speech). While each of these factors may have influenced learners' choices when confronted with multiple forms, Bayley questions the likelihood that a single factor is responsible for these choices. Thus, caution must be observed when identifying factors influencing variation, such as those listed in Table 2-1, as being singularly responsible for the choice of different language forms. Young and Bayley (1996) describe this phenomenon as the principle of multiple causes; it is not a question of determining which single factor is associated with variation but rather the relative strength of the different factors.

Second language research in the variationist tradition assumes that learners' choices are likewise influenced by multiple factors in the choice of a given variant at a given moment in discourse. Preston (1996) likens these variants to a weighted coin; each factor that is determined to influence the choice of one of these variants contributes to the overall probability for one of the variants to be uttered. Thus, a particular variant will not occur randomly; rather, like a weighted coin, it will occur according to the specific weights that represent the influence of linguistic and extralinguistic factors. Preston (2004) explains that speakers use a "sociocultural selection device"

which guides their choice of variants, all of which have been licensed by their grammar. It should not be surprising that L2 learners can also license multiple forms in their interlanguage grammar. As Bayley (2005: 3) adds, "To attempt to explain interlanguage variation as a result of a single factor is to ignore the complexities of SLA." In cases of variation where the learner is confronted with multiple means for expressing the same message (Type II variation), researchers should therefore consider a variety of factors, both linguistic and extralinguistic (or social), in order to carry out a variationist analysis which determines the relative strength of each of the factors associated with variation among multiple forms.

Coveney (2002) provides additional basic criteria for data interpreted under a variationist analysis. First, such an analysis is only useful for explaining the distribution of variants in contexts that allow variation (for example, a variationist analysis of the use of *tu* or *vous* between two close friends will not reveal any useful data because *tu* is likely to be used 100% of the time), though explanations for categorical selection of a particular variant must also be included in any variationist analysis. A more complex issue arises in identifying which forms are, in fact, in variation. Coveney argues that the criterion for identifying a sociolinguistic variable is only fulfilled when all variants carry out the same communicative function; that is, there is "equivalence" across forms. Thus, caution must be exercised in determining whether each variant accomplishes the same communicative function and whether more than one variant is possible in each occurrence.

Once the appropriate contexts for variable forms and all possible variants of this variable have been identified, it is necessary to identify the factors (both linguistic and extralinguistic) that may favor the production of one variant over another and code for each occurrence (or token) of the variable for all of these factors. Though the ultimate goal is to identify all factors that might

significantly impact the choice of one variant over another, one concern of the variationist approach is that we cannot be certain that any analysis has identified *all* factors presumed to account for variation in the data. Nevertheless, the large body of previous research reflected in the factors from Table 2-1 is a good starting point for determining which extralinguistic factors may influence the production of the variable under study.

It must also be noted that, when accounting for the impact of various factors on the selection of a variant, there is a primacy of linguistic factors over social factors (see Preston, 1991, for a review). Briefly, Preston characterizes this tendency via filters that select a specific variant: linguistic filters are strongest due to universal and language-specific conditions on features, resulting from the language-learning process; social status filters then follow, representing more or less permanent aspects of one's identity (also established during the language learning process); and style filters are weakest, representing the social context in which the variant was uttered—even though, as Preston cautions, misreading the registral environment "may have serious repercussions in terms of how one is valued in the linguistic marketplace" (p. 52). Since linguistic factors are highly dependent on the specific variables under investigation, there is less generalizability of specific linguistic factors across sociolinguistic variables; discussion of these factors for the current study will therefore be considered in the context of the two variables chosen (Chapters 4 and 5).

As introduced earlier in this chapter, studies focusing on variation in French have revealed at least a subset of extralinguistic factors found to influence both L1 and L2 variation for multiple variables, though the relevance of these factors may be more dependent on the characteristics of the task design and knowledge of the community of speakers under investigation. For the current study, I will consider the previously identified factors that will be relevant to the profiles of the

learner groups to be chosen, while adding interlocutor language background as an extralinguistic factor in the variationist analyses to be conducted. To my knowledge, this is the first study under the variationist approach to include interlocutor L1 as one of the possible factors influencing sociolinguistic variation in specific variables. Furthermore, its selection as a significant factor in any variationist analyses would render more robust any observations of significant differences measured across different oral production samples of the same speakers.

2.11 Summary of Interlocutor effects, SLA, and sociolinguistic variation

From my initial observations in Chapter 1 regarding the potential influence of interlocutor effects on language production, the current chapter has traced this inquiry through previous literature relevant to interlocutor effects, second language acquisition, and sociolinguistic variation. I have begun with a discussion of second language acquisition outcomes and the situational factors that may impact measures of sociolinguistic variation in learners. I have outlined the previous literature on interlocutor effects, tracing their influence on speaker behavior through Accommodation Theory and Audience Design, followed by the application of these theories to the specific environment of native and non-native speaker interactions. Research on such interactions involving interlocutor language backgrounds has been largely restricted to classroom-based observations and analyses of grammatical proficiency; the current study will address the dearth of data on naturalistic production and analyses of sociolinguistic performance in learners. I have extended the inquiry to the acquisitional challenges (including possible limits on end-state outcomes) involved in attainment of nativelike sociolinguistic proficiency; since examinations of highly proficient learners are comparatively scant, but valuable for determinations of sociolinguistic outcomes, the current study will include this learner population. I have then reviewed the state of the field in L2 sociolinguistic variation more broadly before focusing on the

influential developments in the literature on sociolinguistic variation in L2 French, where I have highlighted the scant attention drawn to the status of the interlocutor in previous research on sociolinguistic performance, specifically in regard to learners of French. Finally, I have motivated a variationist approach for the current study in order to quantitatively examine the potential impact of the interlocutor effect on sociolinguistic variation. In the next chapter, I will detail how I conducted the current study as a response to the questions raised in this discussion of previous investigations into the interaction of sociolinguistics and second language acquisition.

Chapter 3: Methodology

This chapter begins with the process of developing a research methodology based on outstanding issues raised in the discussion of previous literature in Chapter 2. I begin with a discussion on the determination of the level of formality observed in an oral production task between two speakers; since most L2 speakers learn their L2 in a classroom setting, it is particularly interesting to know whether they eventually can use the informal variants in ways that resemble their use by native speakers. This discussion subsequently allows for a formulation of research questions and hypotheses to be addressed in the current study. I follow with a description of the procedure I used for conducting an initial pilot study and a full-scale study based on these questions. I conclude with a discussion of the resulting corpus obtained for this project, including an overview of informal discourse features used by participants in the full-scale study, and a motivation for the selection of specific sociolinguistic variables for testing the influence of the interlocutor effect on informal discourse, to be analyzed in Chapters 4 and 5.

3.1 Informal discourse

Sociolinguistic variables to be examined in a given conversation are often conditioned by the level of formality assumed by the conversation participants. This level of formality may be considered as falling at a certain point on a formal-informal continuum (rather than simply formal versus informal, or written versus spoken), in which the differences between points on this continuum may determined by features of informal discourse (Biber, 1995). Note that each participant's point on this continuum may not necessarily be located on the same point as her interlocutor(s), nor is this location necessarily a conscious choice on the part of each speaker. For most low-proficiency L2 speakers, this determination of formality tends to occur in the rather

artificial language environment of the classroom, where the level of formality may be imposed by the instructor. Potentially concerned by acquisition of "correct" or "proper" forms of the language, or where the focus is on grammatical accuracy, primacy tends to be given to formal or standard forms, even when not stylistically appropriate for the communication task at hand (cf. Etienne & Sax, 2009). When examining acquisition of sociolinguistic variables, the focus of previous research has therefore tended toward learners' non-targetlike use of informal variants when faced with an informal communicative task. However, some researchers have specifically analyzed learner production of formal variants, such as Sax (2003). Sax manipulated the elicitation context to include role-plays at both ends of the formal-informal continuum, with interlocutors instructed not only to assume a formal or informal speaking style, but also to match their dress and general comportment with the appropriate style. Sax found that this difference in formality was a significant factor in the variation produced for all four sociolinguistic variables under study. An analysis of a learner's full sociolinguistic profile would not be complete without measures of production in a formal communicative setting, which may be particularly relevant for language learners who have spent more time in informal settings, such as heritage language learners. However, since the focus of the current study is on adult second language learners of French, most of these learners have likely spent at least part of the acquisition process in a classroom setting, and sensitivity to sociolinguistic variation would be demonstrated by the use of informal forms in stylistically appropriate settings, such as an oral production task in which the learner is asked to have a casual conversation with an interlocutor.

Thus, the type of conversations to be examined for such sociolinguistic variation should show evidence that the speaker treated the conversation as lying toward the informal end of the formality continuum (whether or not the researcher gave any instructions, implicit or explicit,

concerning the type of conversation, and whether or not the speaker followed these instructions). At the informal end of the continuum, one is also more likely to observe authentic communication; indeed, informal, face-to-face conversation is, as Fonseca-Greber and Waugh (2002: 101) characterize it, "the fundamental type of linguistic communication" (see also Chafe (1994)). Therefore, the elicitation procedure should involve a face-to-face communicative context where the learner is aware of the casual nature of the interaction, and where the learner's recognition of this informality can be detected from specific features of her language.

French is abundant in linguistic features appearing in this type of informal communication. Donaldson (2008) provides a list of features typical of spoken French as evidenced by a large body of research (Table 3-1). Production of several such features would indicate a more informal style, though note that not all features need to be present for an informal style to be inferred.

Table 3-1. Characteristic features of spoken French (reproduced from Donaldson, 2008: 50)

Informal feature	**Spoken French example**	**Standard French example**
Ne-deletion	*ils ont pas envie de le faire* 'they don't want to do it'	*ils n'ont pas envie de le faire*
Interrogatives	*pourquoi elle prend le cours?* 'why's she taking the class?'	*pourquoi est-ce qu'elle prend le cours?*
Truncation	*pour le devoir de morpho* 'for the morphology homework'	*pour le devoir de morphologie*
Pronoun reduction	*t'sais l'extrait…* 'you know the extract'	*tu sais l'extrait*
/l/-deletion	*s'i y en a plus de quinze* 'if there are more than fifteen'	*s'il y en a plus de quinze*
Object drop	*mais ils utilisent pas quoi* 'but they don't use (it)'	*mais ils ne l'utilisent pas*
Enfin particle	*enfin bon elle est super* 'so well she's super'	*donc elle est superbe*
Hein particle	*c'est pas mal, hein ?* 'it's not bad, eh?'	*ce n'est pas mal, n'est-ce pas ?*
On for *nous*	*on était trente-quatre comme ça* 'we were thirty-four like that'	*nous étions trente-quatre comme ça*
Vocabulary	*les horaires c'est chiant* 'the hours are crappy'[10]	*les horaires sont mauvais*

This table is not meant to be an exhaustive list of informal features in French. Donaldson (2011a, 2011b) has demonstrated that left- and right-dislocation are also features of informal French, as is subject doubling (e.g., Nadasdi, 1995b), an example of which can be seen in Donaldson's example for informal vocabulary (*les horaires c'est chiant*). Furthermore, other informal discourse particles can be considered in addition to *enfin* and *hein*, such as *quoi*, *genre*, and *machin* (Chanet, 2001; Fleischman & Yaguello, 2004; Mihatsch, 2006) which are used in specific discursive contexts as well; still other particles appear in regional varieties of French, such as *comme* in the *chiac* variety of Acadian French (Chevalier, 2001) and in Manitoban French (Hennecke, 2017), and *comme, fait que,* and *là* for Montréal French (Sankoff, Thibault, Nagy, Blondeau, Fonollosa, & Gagnon, 1997).

[10] All examples here are reproduced from Donaldson's original table; a more accurate translation of this expression would be "schedules are a pain."

If a native speaker uses many (if not all) of these features in a given conversation, then one can presume that this speaker has adopted an informal style; one could presume likewise for an L2 speaker who has learned French in a typical classroom environment. Thus, one can examine the distribution and frequency of these features by each speaker in this informal context (alternatively, one can compare this distribution to the distribution of features in a speaking task that is designed to be formal). This is more straightforward for some variables, due to their binary structure (e.g., the presence or absence of a particular variant, such as the *ne* particle), while other variables have multiple variants that can be placed on a continuum of formality (e.g., interrogative structures), including the possibility of using no overt marker without changing the meaning of their utterance.

In a conversation that is putatively on the informal end of a formality continuum, native and proficient non-native French speakers alike would likely expect the appearance of certain features of informal discourse. The frequency of appearance of each informal variant may be conditioned by a variety of linguistic factors (e.g., phonological environment) as well as sociolinguistic and sociostylistic factors (e.g., the (perceived) level of informality, the topics of conversation, and individual differences in the speakers). At one extreme of this distribution, the informal variants of each feature would appear in each variable context during the course of the conversation (note, though, that for features such as the pragmatic particles *enfin* and *hein*, there is not necessarily an equivalent more formal variant that must appear in its place). Most likely, an informal conversation will contain some more formal, or at least neutral, variants for a given feature. Even in the most informal conversation, it would not be unusual, for example, for a speaker to use the more neutral *voiture* (‘car’) rather than, or in addition to, *bagnole, caisse,* etc. On the other hand, in a given informal conversation, depending on the topics covered, one may not have the occasion to use, for example, object drop, which is favored with verbs of preference. Thus, one

can argue that object drop is a possible, but not obligatory, feature whose appearance can signal an informal conversation. However, due to their relatively high frequency of occurrence, negation contexts are likely to require a speaker to make numerous decisions on retaining or omitting *ne*—whether or not each of these decisions is made consciously by the speaker, and whether or not the speaker is aware of the profile that she is constructing for each variable as a conversation progresses. Nevertheless, regarding *ne*, it would be highly unusual for a proficient speaker, especially a native speaker, to retain *ne* categorically in an informal conversation. Thus, the co-occurrence of certain features, and the frequency of use of informal variants of other features, all contribute to the level of (in)formality that each speaker projects during the conversation.

Beyond the personal and social characteristics of a conversation partner favoring an informal style (as in Sax, 2003), it may be particularly important to learners that the interlocutor's distribution of these linguistic features in a given conversation indicate a more informal style, so that the entire communicative context is favorable to the production of informal variants in learners (and thus greater potential for more nativelike speech production). If this situation obtains, then we can be reasonably certain that any other sociolinguistic variable would be "filtered" through this informal context, even though the learner herself may not show sensitivity to variation for each variable. We can then examine how L2 speakers at different proficiency levels use these variables compared with native speakers in the informal context, and we can compare how these variables are used across multiple speaking tasks when the interlocutor language status is manipulated.

3.2 Research questions

We can now begin defining the scope of an inquiry into potential effects of the interlocutor's language background on learner discourse. Previous studies, based largely on

observation of classroom behavior, indicate that learner attitudes may be influenced by the native language status of their interlocutor. We also know that NNS-NNS interactions may be influenced by issues of identity and divergence. If motivated learners only substantially develop sociolinguistic ability in authentic interaction with members of the target language communities, we may question how learners in these communities interact with other speakers (both native and non-native). Furthermore, if we create an oral production task designed to be on the informal end of the formal-informal continuum (and if stylistically conditioned features observed in the speakers clearly indicate the adoption of a more informal style), then we can empirically observe the extent to which learners use the appropriate variants of sociolinguistically conditioned variables in an environment optimally approximating informal interactions in the target language community. We can then determine whether the interlocutor's language status has a detectable effect on the use of these variables (i.e., the use of the informal variant in an informal speaking task).

The only study to date focusing on interlocutor language backgrounds in French (Dewaele, 2004a) was limited to one variable and one learner population, with a between-group statistical analysis. It would thus be instructive to expand the scope of this analysis, extending the inquiry to 1) observe the same speakers across different interlocutor situations; 2) include a range of learner proficiency levels in order to identify at which level(s) learners might be susceptible to this interlocutor difference; and 3) select and analyze multiple sociolinguistic variables in order to identify whether any interlocutor effects can be detected in more than one sociolinguistic context (with an ultimate goal, beyond the scope of this initial project, of identifying all variables that may be susceptible to this effect).

At this point, I can condense my inquiry to two major research questions: 1) What kind of sociolinguistic ability can be expected from learners of French at various proficiency levels; and 2) How much does the interlocutor play a role in influencing this performance in learners?

3.3 Hypotheses

Concerning question #1, based on previous studies, we can predict that intermediate to advanced learners (such as those in Dewaele (2004a)) will demonstrate greater adherence to standard (or classroom) norms than near-native speakers with regard to sociolinguistic variables. Near-native speakers are expected to pattern closer to native speaker norms (as Donaldson (2008, 20011a, 2011b) has found for dislocated structures), though this learner group may still show non-targetlike deviation.

As for question #2, lower-level learners will be expected to show convergence toward their interlocutor in usage of sociolinguistic variables, as Dewaele (2004a) found with *ne*-retention. Near-native speakers, who may be motivated to use these variables in the same way as NSs do, are expected to be less influenced by the native versus non-native status of the interlocutor, though they may only be able to exhibit targetlike sociolinguistic behavior in interaction with NSs, possibly revealing a secondary interlocutor effect.

3.4 Pilot study

To test these questions, I carried out an empirical pilot study that specifically manipulated the interlocutor language background variable in an informal speaking context. As the research questions are rather specific in terms of the speaker populations and types of interactions needed to test the hypotheses, I generated a new corpus of oral production data, not only to avoid possible misinterpretations of the data in existing NS-NNS corpora (cf. Dewaele, 2007: 18), but also to

address the dearth of data available on NNS-NNS interactions, including interactions between near-native speakers (hereafter abbreviated to *Near-NS*, in order to distinguish from other non-native speakers (NNS)). Since these questions involve Near-NSs in informal interactions, the data elicitation method conceived for the pilot study and the large-scale study was largely informed by work on end-state learners in Donaldson (2008), with audio recordings of spontaneous, non-directed conversation involving dyads between native and near-native speakers. Donaldson found that measures of near-native competence were more robust in L2 French speakers living in France than near-natives living in non-Francophone communities. Thus, the first step in creating a new corpus was to identify near-native speakers living in France.

Through networking with contacts in the United States and France, I recruited several near-native speakers of French for a pilot study conducted in May 2013. In the next subsection, I will report on the main findings from this pilot study as they relate to the development of the large-scale, current study. Full details on the participants, methodology, and results of the pilot study can be found in Appendix A.

3.4.1 Summary of pilot study and results

Ultimately, I was able to recruit six near-native speakers of French (minimum of five years residency in France) to participate in all tasks for the pilot study. These Near-NSs each identified a native and a near-native speaker (friend, spouse, or other acquaintance) with whom they could carry out an informal conversation in French. Each Near-NS was audio recorded in dyadic conversation with their near-native and NS interlocutor, with both conversations lasting approximately 45 minutes each. All participants also completed written questionnaires on their language background and language habits. Each Near-NS also completed an Acceptability

Judgment Task (AJT) as a measure of proficiency (see section 3.5.2.6 for details on this task; see Appendix A for results from this pilot study).

Since Dewaele's study (2004a) of different interlocutor types focused on the *ne*-retention variable in French, the results reported for this pilot study will focus on this variable. The six Near-NSs produced 625 verbal negation contexts with 160 tokens of *ne*, an overall *ne*-retention rate of 25.6%, while NS interlocutors produced 34 tokens of *ne* in 265 verbal negation contexts for an overall retention rate of 12.8%. The between-group difference between Near-NS and NS *ne*-retention was highly significant ($\chi^2(1) = 18.3$, $p < .0001$). For Near-NSs, there was little evidence of a uniform correspondence between *ne*-retention rates and AJT results or the language security index.

Concerning the potential effect of the interlocutor, Table 3-2 reports the *ne*-retention results from the pilot study in comparison with Dewaele's (2004a) *ne*-retention data. Note that the overall results in the first column also include data from a confederate near-native speaker who served as interlocutor for two Near-NSs unable to recruit their own interlocutor (see Appendix A for details).

Table 3-2. Ne-retention results in pilot study and comparison with Dewaele (2004a)

	Overall results (all speakers)			**L1 and L2 results divided by L1 status**			**L2 speakers: results across interlocutor type**		
	ne	Total neg	% *ne*	French status	*ne* / total neg	% *ne*	Interlocutor type	*ne* / total neg	% *ne*
Dewaele (2004a)	628	991	63.4	L1	N/A	36.3	L1 French	N/A	53.5
				L2	N/A	72.8	L2 French	N/A	75.5
Pilot: Near-NSs	199	986	20.2	L1	34/265	12.8	L1 French	93/320	29.1
				L2	160/625	25.6	L2 French	67/305	22.0

In his study, Dewaele (2004a) found that L2 *ne*-retention decreased in conversation with NSs compared to other NNSs, by a considerable margin (53.5% versus 75.5%, a statistically significant difference). In the current pilot study, I found the reverse trend: Near-NSs deleted *ne*

more in conversation with other Near-NSs (22.0%) than with NSs (29.1%). This difference of 7.1% was also statistically significant, though marginally so, at <.05 ($\chi^2(1) = 4.13$, $p = .042$).

Based on the results from these data, there does appear to be a slight effect of interlocutor language background, though in the opposite direction from the effect found in Dewaele (2004a) for *ne*-retention. It can be instructive, then to expand the scope of inquiry to a large-scale study and determine if this interlocutor effect holds in a larger sample of Near-NSs, as well as in learners at proficiency levels more similar to those observed in Dewaele (2004a).

3.4.2 From the pilot to the current study

Despite the highly informal speaking situation that can be facilitated when participants recruit their own conversation partners, I ultimately deemed this procedure to have two limitations concerning the examination of interlocutor effects. First, while pilot study participants were generally able to recruit a native speaker interlocutor, I found that these participants had much more difficulty recruiting another near-native speaker, necessitating the near-native confederate (another interlocutor recruited by me) for some conversations. The logistics of arranging a suitable time and place convenient for all of these participants are not insignificant details. Due to the difficulties in recruiting Near-NSs, each near-native recruited by my participants was also automatically considered a "participant"; that is, their conversation data was analyzed in the near-native "participant" group rather than as a separate "interlocutor" group. Finally, there are bound to be differences in formality between, for example, a Near-NS/NS spousal pairing and a Near-NS/NS friend or colleague pairing (as was the case for two conversations), rendering it difficult to obtain comparable relationships for all pairings between the participants and their two interlocutors. Such dynamics may well account for some of the behavior regarding *ne*-retention observed in this pilot study.

3.4.3 Informing the methodology of the current study

Based on these limitations, as well as other methodological considerations potentially impacting the results obtained in the pilot study (as described in further detail in Appendix A), I revised my methodology in anticipation of a larger-scale study, which was made possible when the following year I was able to return to France for an extended stay. In addition to making logistical changes from the pilot study (e.g., arranging to conduct the study at a more neutral site, rather than meeting some participants in their home), I made one significant change to the procedure: I recruited a native and a near-native interlocutor for each learner group. Though this decision potentially increases the formality of the interaction (since each dyad would then consist of speakers who did not know each other beforehand), it allows for the same level of familiarity for all speakers in each interaction, and it minimizes differences in demographic characteristics across interlocutors. Controlling the choice of interlocutor also reduces attention to the idea that interlocutors of different language backgrounds are the focus of the study (explaining to the participants that they were to recruit a native and non-native speaking partner may have made the interlocutor background feature obvious to the pilot study participants). In addition, recruitment of interlocutors by me eliminates the participants' logistical burden of recruiting other participants to serve as native and near-native interlocutors, which had resulted in the exclusion of several possible Near-NS participants for the pilot study. Finally, in the statistical analysis, the recruited interlocutors' data can be grouped together as "near-native interlocutors" and "native interlocutors" for comparison with the Near-NS participants' data.[11]

Living in France furthermore allowed me to identify multiple possible groups of learners for participation in a full-scale study. To provide a comparison with learners at a proficiency level

[11] Henceforth, near-native speakers recruited specifically to serve as interlocutors will be referred to as "near-native interlocutors," in order to distinguish them from the main groups of highly proficient participants ("Near-NSs").

more similar to Dewaele's (2004a) learners (whose *ne*-retention percentages showed a larger effect for interlocutor L1 than in my pilot Near-NS group), and to expand the scope of proficiency from the pilot study, I recruited a group of L2 French study-abroad learners (referenced hereafter as SA learners) where I was teaching at the Université de Pau et des Pays de l'Adour (hereafter abbreviated to Université de Pau or UPPA) in Pau, France. This university hosts study-abroad students during summer sessions lasting several weeks, and many of these students are placed with host families during their stay. As many studies (e.g., Dewaele, 2004a; Nagy et al., 2003; Regan, 1996) have found, successful L2 acquisition is more likely to be demonstrated in learners who are living in the target language community and who have frequent contact with native speakers. These study-abroad learners satisfy both criteria. Though I would be observing these speakers after a short-term study abroad (about 6 weeks), the daily contact with their host families, as well as daily classes in French and frequent community excursions conducted in French, partially overcome the compromise of sampling learners during a shorter-term study abroad than in longer-term study abroad (where learners often have large individual differences in their contact with native speakers during their stay).[12]

There has been considerable attention in the literature on the length of study abroad (short-term versus longer-term) and possible linguistic benefits to learners, as well as the type of study abroad (homestays versus non-homestays). Some studies find measurable gains in L2 proficiency after a few weeks abroad (e.g., Allen & Herron, 2003; Arnett, 2013; Cubillos, 2013; Cubillos, Chieffo, & Fan, 2008; Hernández, 2016), though others have questioned the linguistic benefits of such programs (e.g., Davidson, 2007, 2010; Freed, 1990; Wilkinson, 1998, 2002). Magnan and

[12] Before recruiting these short-term study-abroad students, I initiated recruitment of other learners who were participating in either a semester- or year-long study-abroad program at the Université de Pau, but who were not participating in a homestay. Ultimately, I was only able to recruit three speakers, and only one conversation task was carried out for each speaker. The data from these conversations are not included in the current study.

Back (2007) found no relationship between target language use and speaking gains, noting that most of their study-abroad participants did not form social relationships with native speakers in the community. Castañeda and Zirger (2011) lament that short-term study abroad has inherently limited benefits due to the brief time that participants are able to develop social relationships with native speakers; the authors thus stress the importance of developing short-term programs that maximize participant involvement in the native speaker community. Kennedy Terry (2017) also found that a longer length of time in the native speaker community (academic year versus one semester) positively correlated with targetlike patterns of variation in French /l/-deletion, though this length of time was not selected as a significant factor in her variationist analysis. Martinsen (2010) noted that oral proficiency increased the most in short-term study-abroad students who had the highest scores on a cultural sensitivity assessment, suggesting that these learners were already primed to profit the most from a brief stay in the target language community; Reynolds-Case (2013) also found gains in cultural and pragmatic competence in a short-term study abroad. As for study-abroad homestays with native speakers, many studies report that this type of sojourn does not yield significant proficiency gains for learners over non-homestay study abroad (e.g., Martinsen, 2010; Rivers, 1998; Segalowitz & Freed, 2004; Vande berg, Connor-Linton, & Paige, 2009); however, there is correlation between student satisfaction in the homestay and oral proficiency gains (Di Silvio, Donovan, & Malone, 2014; Hardison, 2014; Schmidt-Rinehart & Knight, 2004). In short-term homestays, learners and hosts may also be more likely to maintain meaningful interaction given the brief time period for cultural and linguistic exchange; Pryde (2014) found that over the course of an 11-month sojourn, opportunities for meaningful interaction between learners and hosts at the dinner table decreased over time. Hardison (2014) also points to the quality and depth of interaction with NSs as important considerations over simply the quantity

of learner-NS interaction. Furthermore, Kinginger and Carnine (2019) draw upon case studies to highlight the different outcomes that different types of host families can have on learners (e.g., "empty nesters" modeling a more pedagogical version of the target language versus families with children who expose the learner to multiple informal styles). Concerning sociolinguistic competence, the evidence seems to be stronger that homestays increase the use of stylistically appropriate informal variants (cf. Rehner et al., 2003; Uritescu, Mougeon, & Handouleh, 2002). At the very least, then, the group of learners that I was able to recruit—participating in a short-term study-abroad program with homestays—is an increasingly frequent type of study-abroad learner group analyzed in SLA research, and is likely to be suitable for an analysis of awareness of sociolinguistic norms potentially influenced by interlocutor characteristics.

On a dyadic level, I hypothesized that this group of SA learners would interact with the near-native speaker in the same way as with the native speaker regarding sociolinguistic variation such as *ne*-retention; that is, there would be convergence in both conversations, regardless of the interlocutor L1 differences. Therefore, I decided to include a third conversation between two SA learners, in order to examine whether any interlocutor effects on sociolinguistic variation could be attributed to the interlocutor's (perceived) level of proficiency in addition to their native language status. Thus, the oral production task for this group of learners would involve three conversations: SA learner with native speaker (L1 French); SA learner with near-native speaker (L2 French); SA learner with SA learner (L2 French). This arrangement is similar to the dyads in Porter (1986).

Given the methodological changes described above, I determined it necessary to recruit a new group of near-native participants who had not participated in the pilot study. Due to further networking with friends and colleagues, I had a larger base of potential near-native speakers from whom to recruit participants for the current study, with at least 10 speakers available at two

different sites in France. This allowed for a subsequent methodological change addressing a fundamental issue arising with an examination of interlocutor L1 effects. That is, are any differences in variation attributable solely to the immutable characteristic of the interlocutor's language background (as L1 or L2 speaker of French), or is an effect on variation detectable based on the actual native or non-nativelike speech produced by the interlocutor during her interaction with the Near-NS participant? In other words, is the Near-NS's *perception* that an interlocutor has native or non-native status the driving factor behind interlocutor effects, or must this effect be reinforced by the interlocutor's actual speech patterns? While learners at lower proficiency levels may not perceive these differences in native versus nativelike speech, highly proficient learners may be susceptible to what can be considered "frequent reminders" that their interlocutor is not a native speaker, due to non-targetlike syntax, phonology/intonation, or vocabulary. Such reminders of non-native linguistic competence may impact Near-NSs' use of variable stylistic features in French more so than the presentation of the interlocutor's identity as native or nonnative.

Teasing out the possible effect of the interlocutor's linguistic competence involves two groups of interlocutors. The first group, testing the perception of native versus non-native status, requires recruitment of interlocutors assuming an identity as a native or as a non-native speaker in a given conversation. It would be feasible to have an interlocutor introduce herself as a non-native speaker and then subsequently demonstrate nativelike speech (rather than the other way around), since there is evidence that certain highly proficient L2 speakers can pass as native speakers even when rated by native speaker judges (Abu-Rabia & Kehat, 2004; Ioup, Boustagui, El Tigi, & Moselle, 1994). Ideally, the same person would then serve as an interlocutor assuming the identity of a native speaker, but for a conversation with a different Near-NS—and, ideally, this interlocutor

would produce the same general speech patterns across both conversations, with the presentation of the L1 identity being the only variable manipulated.

Such a task can be accomplished with bilingual speakers, defined as those who can pass for native speakers of both languages (whether simultaneous or sequential bilinguals). In the context of speaking with native English learners of French (the Near-NSs), this arrangement involves recruiting interlocutors who are English-French bilinguals adopting either an L1 English or an L1 French identity for each conversation. The caveat is that such interactions may require the bilingual to "steer" the topics of conversation where necessary (or at least withhold certain details) in order to speak about aspects of their personal background that are consistent with the adoption of one or the other identity.

The second group of Near-NSs, testing the native versus non-native status reinforced by the interlocutors' actual speech patterns, carry out conversations with a "control" pair of interlocutors reflecting a more "traditional" set of dyads as in the pilot study: one interlocutor who is an authentic native speaker and one interlocutor who is an authentic non-native speaker. However, to ensure that the second Near-NS group is exposed to non-targetlike speech, it would be necessary to recruit a near-native interlocutor demonstrating at least some recognizable deviance from L1 French syntax and phonology.

If differences in the Near-NS' use of sociolinguistic variables obtain in both groups, then this supports the notion that the simple characterization of an interlocutor as native or non-native is sufficient for creating an interlocutor effect on learners, regardless of the interlocutor's actual speech patterns. If differences obtain in the second group but not the first, then there would be evidence that the interlocutor effect is also dependent on the interlocutor's adherence to, or deviance from, native speech patterns.

My network of contacts in France allowed for such a methodology with two groups of Near-NSs at two different sites in France. First, I recruited bilingual English-French students from courses that I was teaching at the Université de Pau, and I established a list of nearly 20 Near-NSs in the local community. I also established a substantial network of contacts in Lille, from which I was able to identify over a dozen potential Near-NS participants, one of whom agreed to serve as the primary near-native interlocutor and who, according to my observation of her L2 speech, would be identifiable by other Near-NSs as a non-native speaker in both syntax and phonology.

Figure 3-1 provides a schematic of the conversation arrangements for the participants at each site, where *n* is the number of expected learners for each site, and where the lines connect the types of participants for each dyadic conversation.

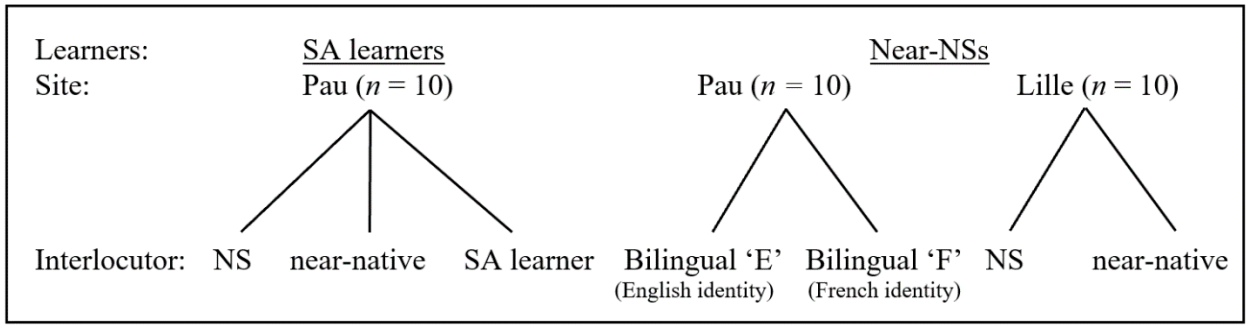

Figure 3-1. Organization of participant groups in the current study

Recall that near-natives recruited specifically to serve as interlocutors are denoted as "near-native" or "near-native speaker" rather than "Near-NS." As Figure 3-1 indicates, these include one of the interlocutors for the SA learners and one interlocutor for the Near-NSs in Lille. To summarize, each SA learner will have a dyadic conversation with a NS, a near-native, and another SA learner. Each Near-NS in Pau will speak with two bilinguals, one adopting an English identity and the other adopting a French identity (with each bilingual adopting an English identity for half of their

total conversations and a French identity for the other half). Finally, each Near-NS in Lille will speak with a NS and a near-native recruited to serve as interlocutor.

3.5 The current study

With the organization of the participant groups established, this section elaborates on the profiles of the participant groups and outlines the procedures administered to each group of learners comprising the current study. Details on the SA learner group are provided first, followed by the two Near-NS groups.

3.5.1 SA learner group

The SA learner group was recruited during May and June 2015, and the study was conducted on the UPPA campus during a two-day period in July 2015. The following subsections detail the characteristics of the participants and their interlocutors, as well as the procedure used for this learner group.

3.5.1.1 Participants

The SA learner group ultimately consisted of eight short-term study-abroad students. The students were all enrolled in a summer French program for international students at the Université de Pau. All students who participated in the study were either living with a host family during their two-month stay in Pau or had lived with a host family during a prior stay in France. Though students from around the world enroll in these intensive French courses, the majority of students come from American universities having partner affiliations with UPPA; all students recruited for the current study were from American universities and all were native speakers of English, with a range of 1-7 years of French study (mean = 4.5 years). The age range was 18-21 (mean = 20).

I determined that the Acceptability Judgment Task administered to Near-NSs would have been too difficult for SA learners and thus ineffectual in measuring their proficiency. Therefore, I administered to all participants a *c*-test modeled from Renaud (2010), adaptable to a range of proficiency levels. In this timed (10-minute), written test, participants are given two paragraphs of text in French, in which every other word has the latter half of its letters replaced by blanks, and the participants are instructed to complete each of these words (n = 50) in a logical manner according to the context of the passage (see Appendix D for the test and expected answers). On this task, SA learner scores ranged from 30-46 (mean = 38); according to previous measures in Renaud (2010), this places all participants in a pre-advanced group (all scored higher than the mean score of a group of fourth semester students, but none scored higher than the mean of an advanced group). Thus, the *c*-test indicates that this group is more homogeneous than the apparent large variation in terms of years of French study (range = 1-7) would have otherwise indicated. Indeed, in general, all SA learners were capable conversation partners; they rarely produced long pauses or incomprehensible speech, even though an informal review of their recorded speech revealed a considerable number of non-targetlike grammatical errors.

3.5.1.2 Conversation partners (interlocutors)

I recruited two female speakers of French, one a native speaker and the other a near-native speaker, to serve as conversation partners. The native speaker, age 30, was born and raised in the south of France; though she also reported proficiency in English and Arabic, she spoke only in French with the learner group. This speaker, identified in the data as 'SoF,' works at a middle school near Pau in a role similar to an educational paraprofessional, assisting a vision-impaired history/geography teacher in the classroom.

The near-native speaker (L1 American English) was chosen due to her high proficiency in French. This speaker, identified as 'AmE,' is the confederate (speaker 1A) in the pilot study discussed in section 3.4 and in Appendix A. Her updated characteristics for the current study are listed below.

ID	Age	Sex	CoB	AOI	AOE	AOA	LOR	Educ.	Profession
AmE	27	F	USA	14	18	18	4.5	MA	student/ *lectrice*

Abbreviations: CoB = country of birth; AOI = age of first instruction in French; AOE = age of first major exposure to native speakers of French; AOA = age of continuous exposure to French (beginning of long-term stay in France); LOR = length of residency (years); Educ. = highest education level completed

At the time of the data collection with SA learners, this speaker was working as a *lectrice* (English instructor) at Université Lille 3 and concurrently pursuing a doctoral dissertation in French Linguistics.

In interactions with the students, both interlocutors were dressed casually and adopted an informal conversational tone. Informal post-hoc observations of these speakers' oral production data did not reveal any apparent "overaccommodation" or "foreigner talk" that may appear in NS-NNS conversation and in other high-status/low-status interactions (cf. Zuengler, 1991). Both speakers were unaware of the specific nature of the study (that is, the sociolinguistic variables to be analyzed for possible interlocutor effects). Furthermore, the status of these interlocutors as native/non-native speakers was not explicitly mentioned to the SA learners before they began the conversation tasks. Though speaker AmE was present for initial recruitment of some of the SA learners, she was not explicitly identified as a native speaker of English or French, nor was she identified as an investigator in the study.

As in Porter (1986), the third interlocutor consisted of another SA learner. To maintain complete consistency with the characteristics of the previously described dyads, it would have been ideal to recruit a learner at the study-abroad level to serve as one interlocutor for all other

learners. However, this was logistically impossible, given the time constraints on this learner group. Therefore, I created dyads with the pool of eight SA learners recruited. On one hand, these speakers were familiar with each other, since they were taking the same classes together; on the other hand, this situation created a context in which an additional aspect of the research questions could be addressed; that is, whether learners would be more influenced by the familiarity dynamic (and possibly increase their use of informal sociolinguistic variants), or whether they would be more influenced by the interlocutor language background, creating a more artificial classroom-like dynamic (and decreasing their use of informal variants).

3.5.1.3 Procedure

To begin the procedure, I invited each participant in groups of two to my office at the UPPA campus. The SA learners reviewed the informed consent statement and provided verbal consent to participate. They then completed the *c*-test in my presence. Afterward, I led each learner individually to another room where they met either the native or the near-native interlocutor; I then turned on the recorders and left the room. One learner began in a NS-NNS dyad with the native French speaker, and the other began in a NNS-NNS dyad with the near-native speaker, switching conversation partners for the second conversation. The final audio recording was then carried out between the two SA learners. Each conversation generally lasted between 20-25 minutes, with only the two speakers present. All participants were provided with snacks and drinks in these classrooms which contained chairs seated around long conference tables. Participants could take a break between each conversation if they wished.

For each conversation, as in the pilot study, I instructed the learners to chat with their partners on any topic they wished. However, due to the lower proficiency levels of the learners, written topic prompts were provided for all conversations, in case the speakers could not sustain

conversations on their own spontaneously chosen topics (see Appendix C for a list of these topic prompts). The prompts were chosen for their potential to facilitate extended discourse on a casual topic (e.g., a typical day in their summer session, places they like to go in Pau, travels in France/Europe, what they like/dislike about French food), and different topics were suggested for each type of dyad, in order to avoid intentionally providing learners with the same topic across multiple conversations (though they were not instructed to avoid repeating topics from conversation to conversation). No other specific topic manipulation was made as previous sociolinguistics studies have done (e.g., interviewing participants about pre-selected serious versus casual topics). The prompts were also provided in written form as noun clauses or sentence fragments rather than as questions, in order to allow participants to begin speaking on their own or to invite their partner to speak without restricting the type of interrogative structure that participants might use to (re-)initiate conversation. Importantly, these written prompts were placed between the two speakers for each conversation, in order to avoid as much as possible any notion that speakers were expected to adopt a specific conversational role (e.g., one speaker holding a list of questions and acting as interviewer), or that the native/near-native interlocutors had any pre-assigned role such as an interviewer.

Afterward, I invited the learners back to my office to complete a background questionnaire on language use and cultural integration. Finally, each learner met separately with me for a recorded debriefing session in English.

3.5.1.4 Results: Language security index

In addition to the *c*-test measure for proficiency, I also determined a self-reported measure of proficiency based on the learners' responses in the language background questionnaire, specifically those concerning their abilities to read, write, speak, and hear French, as well as ability

and motivation to pass as a native speaker (questions #17-22 and #24; see Appendix E). Each question had five numbered responses, with 1 reflecting the lowest and 5 reflecting the highest ability and motivation. The responses for each speaker were averaged to produce a "language security index," reported for each speaker in Table 3-3.

Table 3-3. Measure of language security in SA learners

Speaker ID	**Abilities in French**					**Motivations in French**		**Security Index (1-5)**
	Read	Hear	Write	Speak	Accent rating	Pass as NS	Attempt to pass as NS	
1S	4	4	3.5	4	3	3	5	3.8
2S	4	4	4	4	4	3	2	3.6
3S	4	3	3	3	3	1	1	2.6
4S	4	4	4	4	3	3	2	3.4
5S	5	4	4	4	3	1	4	3.6
6S	4	4	3	3	3	1	4	3.1
7S	4	3	3	3	3	1	3	2.9
8S	4	4	4	3	3	2	4	3.4
Average	4.1	3.8	3.6	3.5	3.1	1.9	3.1	3.3

Whereas the *c*-test may be a good measure of written fluency (and attention to written forms) as a proxy for overall proficiency, a measure of proficiency (albeit self-reported) obtained by this language security index may also capture other factors involved in a speaker's overall competence and motivation for speaking and learning the language. Such motivations may also influence the speaker's awareness of sociolinguistic variation.

3.5.2 Near-NS groups

I recruited both Near-NS groups during the spring of 2016. In Pau, I conducted the study on the UPPA campus over several sessions in June 2016. In Lille, I conducted the study mainly at the Université de Lille 3 campus (see section 3.5.2.5 for more details) over several sessions in June and July 2016. Details on the Near-NS participants and their interlocutors are outlined in the

following subsections. Since the procedure for these groups was modified somewhat from the procedure for the SA learner group, additional details are provided in section 3.5.2.5.

3.5.2.1 Pau

In all, 13 L1 English speakers participated in the study in Pau, including the oral production task (audio-recorded conversations) and all written tasks. During the debriefing, I immediately determined that one putative Near-NS speaker was in fact a bilingual from birth, having grown up in France with L1 English parents; his data is excluded from this analysis. Ultimately, I excluded two more speakers from analysis, due, on the other hand, to considerable deviance from nativelike phonology and syntax. The remaining 10 speakers form the dataset from Pau. Demographic information is provided in Table 3-4 below.

Table 3-4. Near-NS participants: Pau

ID#	Age	Sex	CoB	AOI	AOA	LOR	Educ.	Profession	Partner L1
1P	76	F	UK	11	43	13	Assoc. degree	Retired secretary	English
2P	62	F	UK	11	25	36	BA	Translator/CEO	French
3P	47	F	UK	11	23	24	BA	ESL teacher	French
4P	44	F	USA	13	24	20	BS	Chemical engineer	English
5P	68	F	USA	17	41	27	MA/ MBA	Retired (no profession specified)	French
6P	66	M	UK	11	60	6	MA	Retired French teacher	English
7P	32	F	UK	11	22	10	BA	ESL teacher/ Translator	French
8P[13]	53	F	Canada	8	35	18	BA	ESL teacher	French
9P	38	F	UK	8	29	9	MA	Wine shop owner	French
10P	37	F	USA	14	30	7	MA	ESL teacher	French

Abbreviations: CoB = country of birth; AOI = age of first instruction in French; AOE = age of first major exposure to native speakers of French; AOA = age of continuous exposure to French (beginning of long-term stay in France); LOR = length of residency (years); Educ. = highest education level completed

[13] This participant was born in Québec but grew up in an English-speaking community in Ontario.

The average age of the Pau participants was 52.3, and the average length of residence was 17.0 years. None of the participants in the current study recalled participating in a study similar to Donaldson's, which was carried out in 2006, and which had included several participants from the Pau area. Donaldson (personal communication, March 29, 2019) confirmed that none of the near-native participants in his study participated in the current study.

3.5.2.2 Lille

Initially, I recruited 10 near-natives in Lille. When the original near-native interlocutor recruited for the study could not participate due to scheduling conflicts, I recruited one of the near-native participants to serve in the "near-native interlocutor" role. All nine remaining Near-NSs in Lille who participated in the study were included for analysis; see Table 3-5.

Table 3-5. Near-NS participants: Lille

ID#	Age	Sex	CoB	AOI	AOA	LOR	Educ.	Profession	Partner L1
1L	27	F	Canada	7	22	5	MA	ESL professor	French
2L	30	F	USA	12	24	6	MA	ESL professor	English
3L	30	F	Australia	25	25	5	MA	ESL professor	French
4L	48	M	USA	14	32	18	MA	ESL professor	French
5L	63	F	UK	11	20	43	BA	Retired ESL professor	French
6L	33	M	USA	14	28	5	MA	ESL professor	French
7L[14]	53	M	Ireland	11	24	29	MA	Medical institute teacher	French
8L	29	M	USA	10	24	5	MA	ESL professor; student	N/A
9L	26	F	USA	13	22	5	BA	ESL professor	N/A

The average age of the Lille participants was 37.7, and the average length of residence was 13.4 years. The differences in age and length of residence compared with the Pau group can be partly attributed to the social network from which I recruited participants. That is, nearly all participants in Lille were recruited through contacts with teachers of English at Université Lille 3, most of

[14] This participant lives and works across the border in (French-speaking) Belgium, but regularly visits Lille.

whom were working as *lecteurs/lectrices*, an instructor position often offered to students pursuing graduate degrees. Pau, in contrast, involved more recruitment outside the Université de Pau community, which drew a comparatively older population with longer length of residence (for example, three Pau participants were nationals of the United Kingdom who had retired in Pau).

3.5.2.3 Interlocutors: Pau

In Pau, I initially recruited four bilingual native speakers of French and English (two males, two females). However, one of the females was unable to participate due to scheduling conflicts. The three remaining bilinguals thus included two males (identified as 'Fr' and 'Th') and one female (identified as 'Ch'). All three were students at the Université de Pau. One of the male students (Th) had just finished the first year of the master's program in teacher training for the CAPES exam (certification to teach English at the secondary school level). The other two students had just finished a *licence* degree (equivalent to a bachelor's degree) in English. The remainder of this section details the bilingual status of each of these speakers.

Bilingual Ch was born in England and lived there exclusively until she was about four years old. Subsequently, she began visiting France each summer to spend time with her English grandparents, who had moved to the south of France upon retirement. At the age of 10 she moved to France to live with her grandparents, whereupon she enrolled in the French school system, completing the remainder of her primary school studies, as well as *collège*, *lycée*, and university studies. She returns to England on occasion to visit the rest of her family. Based on informal observations by me, and mentioned by several of the interlocutors in her conversations, she has adopted certain features typical of native French speakers across the south of France (e.g., replacement of nasal vowel /ɛ̃/ with oral vowel /ɛ/ followed by velar nasal /ŋ/ as in *demain* ('tomorrow'); pronunciation of final /s/ in non-liaison contexts for *moins* ('less')).

Bilingual Fr was born in the south of France to a French father (whose parents immigrated from Poland) and an English mother. He has lived his entire life in France and has completed all levels of formal education, including his *licence* degree, in France. He has relatives living in England and occasionally travels there. Informally, one of his interlocutors recognized a local accent (due to pronunciation of final /s/ in *moins*), though Fr's accent appears to contain fewer regional features overall compared with Ch's accent.

Bilingual Th was born in Paris to a father from South Africa and a mother from England. He spent his early childhood in Paris and recalls speaking French during that time. At age four, his family moved to England and spent six years there. At age 10, his family moved to the south of France, where Th completed primary school, *collège*, *lycée*, and university studies. Before resuming his studies at the master's degree level, he spent one year living in Spain; he reports communicative fluency in Spanish. I did not observe Th's French accent to contain any immediately distinctive regional features.

Despite some differences in the backgrounds of each of these individuals, all of these speakers have resided in France since at least their primary school studies, and they satisfy a broad definition of "bilingual" as evidenced by successful acquisition of both English and French as children. As they were enrolled in English courses at UPPA during my stay there, my main interaction with these bilinguals was in English. Though their English-speaking abilities were only important for the purposes of identification as bilingual English-French speakers, all of them spoke dialects of British English without any obvious phonological or syntactic influence from French. Their French-speaking abilities, however, were more pertinent for the purposes of the current study. Upon review of their conversations conducted in French, none of them demonstrated any obvious recurring phonological or syntactic influence from English. However, there were 12

grammatical errors identified in Ch's speech (all but one involving gender agreement of articles and adjectives), seven under her English identity and five under her French identity. There was also one instance of a non-standard preposition in Th's speech under his French identity (with possible influence from English: *sur la télé*, lit. 'on the TV'; compare with standard French *à la télé*, 'on TV'). I detected no grammatical errors in Fr's speech.

Before beginning the study, I instructed these bilinguals on the nature of their role as interlocutors. I explained that they would adopt either an Anglophone or Francophone identity, but that they would speak French during their entire interaction with the other participants. If (and when) they were asked about their backgrounds by the Near-NSs, I advised them to highlight aspects of their English or French identity where appropriate. Otherwise, I reiterated that the goal of the speaking task was simply to have a casual conversation with the other participants, on any topic they wished; I provided no pre-selected topics for any of these conversations. Note, finally, that in my corpus analysis, I added an identifier to each bilingual's speaker ID when it was necessary to indicate the identity that bilingual had adopted for that conversation (that is, 'ChE' for a conversation where Ch adopted her English identity versus 'ChF' for her French identity, as well as FrE versus FrF and ThE versus ThF).

3.5.2.4 Interlocutors: Lille

Before beginning the oral production tasks in Lille, I outlined to each of the recruited interlocutors that the task was simply to have a casual conversation in French. The near-native interlocutor 'SaE' served as the primary near-native conversation partner for the Lille participants. Due to scheduling concerns, she was not available for three of the conversations, so a second near-native interlocutor ('JeE') was also recruited. The background information for these two speakers is as follows:

ID#	Age	Sex	CoB	AOI	AOA	LOR	Educ.	Profession	Partner L1
SaE	36	F	USA	11	29	7	PhD	ESL professor	French
JeE	35	M	USA	14	22	13	BA	ESL professor	N/A

Both speakers displayed some non-nativelike features of syntax, phonology, and intonation. For SaE, there was an average of nine grammatical errors detected per conversation (range: 4-13); JeE also had an average of nine grammatical errors per conversation (range: 6-11). Such characteristics of these individuals' speech would thus indicate to highly proficient speakers that these individuals were non-native (though highly proficient) speakers.

The native speaker, 'CaF,' was recruited through contacts in the community and served as the primary NS interlocutor in Lille. CaF is a 26-year-old female who was born in Lille. Other than living for six years of her childhood in Bordeaux, she has remained in Lille, where she earned a master's degree in public law and now works as a *juriste* (a legal assistant). Due to CaF's work schedule, she was not available for two of the conversations, so I recruited a second native speaker ('KeF') through contacts with the community. KeF, a male, is 25 years old, from Douai (near Lille) in the north of France. He has a master's degree and works as a middle school teacher (history/geography) in the Lille area.

3.5.2.5 Procedure

The procedure for the Near-NS groups was modified slightly from the pilot study and from the SA learner group. To begin, I met with all participants individually. In Pau, I met everyone either in my office or in an adjacent computer lab in the English department on the UPPA campus. In Lille, I met all but two participants on the Université de Lille 3 campus in a teacher's lounge area near the English department; to accommodate the schedules of the other two participants, I met them in a quiet lounge room beneath a bar/café in downtown Lille, offered by the NS

interlocutor, CaF, who served in her spare time there as a volunteer bartender for a non-profit organization.

All participants began by reading the informed consent statement and providing their verbal consent to participate. Next, they completed the timed *c*-test in my presence. To prepare the participants for the conversation tasks, I explained that each Near-NS would have two casual, one-on-one conversations in French, though I did not give any indications about the native status of each interlocutor. I instructed each Near-NS to speak with her interlocutor for at least 30 minutes and to develop the conversation with any topic that came to mind; that is, no conversational topics were prescribed or suggested. This also means that neither speaker was assigned a particular role in this production task, thus minimizing any perceived power imbalance between participants.

I then brought in the first interlocutor and set up the recording equipment. For all locations where the speakers were audio recorded, I attempted to provide a casual speaking environment. In Pau, the speakers conducted these conversations in my office, where I had arranged two lounge chairs next to a large window and a table with snacks and drinks. In Lille, the conversations were recorded in one of three locations: the teacher's lounge for the English teachers at Université de Lille 3; another lounge room provided by the English club on campus; or the lounge room in the basement of the downtown bar/café. At all of these Lille locations I also provided snacks and drinks for the participants in the recording location. The recording equipment was similar to that used in the pilot study: a digital audio recorder connected to a wired lapel microphone was placed near each speaker, who attached the microphone to his or her clothing. When the participants indicated that they were ready to begin the conversation, I checked that the recorders were working before leaving the room.

To minimize ordering effects, in Pau I alternated the order of identities in which the 13 Near-NSs met the bilingual speakers: seven began with a bilingual adopting a French identity, and six began with a bilingual adopting an English identity. I also arranged so that each bilingual conducted half of their conversations (or half minus one for those who participated in an odd number of conversations) under each identity. In Lille, the ordering was more imbalanced due to scheduling conflicts with the interlocutors: three of the Near-NSs began with the native interlocutor and six began with the near-native interlocutor. Each of these conversations lasted between 30-40 minutes, with a short break in between.

After these two conversations, each Near-NS then met with me to conduct a shorter, recorded conversation in French (10-15 minutes). This was followed by the language background questionnaire and, for some participants, the Acceptability Judgment Task, which was identical to the AJT administered in Birdsong (1992) and Donaldson (2008). Other participants with time constraints elected to complete the AJT on their own at a later time (a procedure that Donaldson allowed as well). The debriefing was then conducted and recorded in English. After the debriefing, I then offered all Near-NS participants the choice between 10 euros, a bottle of local wine, or a box of desserts, as a token of appreciation for their time. Due to the increased time commitment, I also compensated native and near-native interlocutors five euros for each conversation in which they participated.

3.5.2.6 Results: Acceptability Judgment Task

In order to obtain a measure of grammatical competence in the Near-NSs, and to allow for a certain level of comparison with previous studies on near-native speakers, I administered the Acceptability Judgment Task (AJT) used in Donaldson (2008) and Birdsong (1992). I followed Donaldson's procedure and replicated Birdsong's (1992) original 76-item judgment task, which

tests participants on their acceptability of several grammatical constructs that pose difficulty for learners of French, including *ce* versus *il* as subject pronoun (e.g., *Marie a dit de Jean que c'est / qu'il est un génie* 'Marie said of Jean that he is a genius'), *en*-avant clitic movement structures (e.g., *Elle a lu ce livre. Elle en aime l'auteur*. 'She read this book. She likes the author [of it].), *that*-trace (e.g., *Que dis-tu que Marie a acheté ?* 'What do you say that Marie bought?'), and middle voice (e.g., *Cette maison s'est vendue d'elle-même* 'This house sold itself'). For each item, consisting of a sentence or a sequence of two sentences, participants are asked to rate its acceptability on a five-point scale A-E, where A is not at all acceptable and E is completely acceptable. They are instructed to answer based on their first intuition, without taking time to think about grammar rules, and they are instructed that there are no correct or incorrect answers, even though about a third of the items were actually rated as ungrammatical by native speakers in Birdsong (1992). As in Donaldson (2008), there was no time limit for completion of this task, and to avoid potential ordering effects, three different randomized versions were used. The entire task, administered all in French, can be found in Appendix F. Nine of the ten Near-NSs retained for analysis in Pau, and seven of the nine Near-NSs in Lille, completed the AJT. The remaining three Near-NSs (speakers 7P, 4L, and 5L) had initially elected to complete the AJT at a later date but did not send me their results after my follow-up requests; due to time constraints, one of the NSs (bilingual Th) was unable to complete the AJT.

I begin an analysis of the AJT results by comparing the average responses on each item according to its grammaticality, as determined by native speakers in Birdsong (1992), and I compare these averages to that of near-natives in previous studies. Note that I the converted judgments on the five-point A-E scale to numerical 1-5 scores, so the higher the score, the more grammatical the item was rated. Table 3-6 compares the average results for the Near-NSs in Pau

and Lille with the Near-NSs in the pilot study and the near-natives in the studies by Birdsong and Donaldson.

Table 3-6. Acceptability judgment task results for Near-NSs across studies

Grammaticality of item	**Pilot study**	**Near-NSs in Pau**	**Near-NSs in Lille**	**Birdsong (1992)**	**Donald-son (2008)**
Grammatical items (*n* = 44)	3.68	3.91	3.65	3.72	3.64
Ungrammatical items (*n* = 26)	2.14	1.98	1.97	1.94	1.77
Questionable grammaticality (n = 7)	2.74	2.60	2.65	2.53	2.23

As this table indicates, for items that are ungrammatical and of questionable grammaticality, Near-NSs in Pau and Lille compare even more closely with Birdsong's and Donaldson's near-natives than those in the pilot study; furthermore, Near-NSs in Pau accept grammatical items at the highest rate of all these groups. Based on these results, then, Near-NSs at both sites fall well within the ranges of previously measured learners demonstrating near-native proficiency, at least with respect to syntactic competence. Furthermore, though I have results for only four NSs in the current study (two of the Pau bilinguals and both NSs in Lille), it can be instructive to compare their average AJT scores to the NSs in the Birdsong and Donaldson studies. Table 3-7 provides these comparisons.

Table 3-7. Acceptability judgment task results for NSs across studies

Grammaticality of item	**NSs in current study (*n* = 4)**	**Birdsong (1992) (*n* = 20)**	**Donaldson (2008) (*n* = 9)**
Grammatical items (*n* = 44)	3.94	3.55	3.90
Ungrammatical items (*n* = 26)	1.79	1.82	1.86
Questionable grammaticality (n = 7)	2.46	2.19	2.41

While my NSs accepted ungrammatical items slightly more than in either of the previous two studies, my NSs nearly match Donaldson's NSs on grammatical items, and ungrammatical items are accepted at extremely similar rates across all three studies, Thus, at least for the small sample of NSs in the current study, their behavior closely matches larger samples of previously examined native speakers.

Birdsong and Donaldson conducted two-tailed *t*-tests on the native and near-native response patterns for each of the 76 items in the task, in order to determine how similarly the two groups matched in a measure of grammatical competence, by comparing the average response rating for the two groups on each item (where the A-E scale was converted to a numerical 1-5 scale). For Birdsong's participants, near-native speaker responses as a whole significantly differed from native speaker responses on 17 of the 76 items, while Donaldson's near-natives differed on 5 of the 76 items. Donaldson thus concluded that his near-natives possessed a level of grammatical competence equal to or superior to Birdsong's near-natives. Since the format of the current study was designed to include only three native speakers, a *t*-test on each item for comparing Near-NSs and NSs in my study would not be statistically valid. However, Donaldson granted me access to the AJT results for each item in his native speaker group, allowing for a statistically valid comparison with my Near-NSs and a NS group of nine speakers. I therefore conducted individual *t*-tests (two-tailed independent samples) for each item and compared each of my Near-NS groups to Donaldson's NS group.

For Near-NSs in Pau ($n = 9$, the same number as Donaldson's NSs), there were significant differences from Donaldson's NSs on 4 of the 76 items, and for Near-NSs in Lille ($n = 7$), there were significant differences on 12 of the 76 items. Based on this measure, the Near-NS group in Pau fares slightly better than Donaldson's near-natives in terms of syntactic competence, while the

Near-NS group in Lille falls between Birdsong's and Donaldson's near-natives on this measure. Moreover, since the most proficient Near-NS in Lille (speaker 4L, based on the fewest grammatical errors) did not complete the AJT, it is possible that the Lille group would have significantly differed from NS results on even fewer items. I can reasonably conclude, as Donaldson did for his near-natives, that each of the two Near-NS groups in my study is comparable to other such populations examined in previous literature, in terms of syntactic competence.

For individual speakers, an additional analysis can be carried out by comparing Near-NS judgments with NS judgments in previous studies in order to determine the extent to which each Near-NS differed from NS averages. I took the average rating of the AJT results on each of the 76 AJT items for NSs reported in Birdsong (1992) and Donaldson (2008) and computed the difference for each speaker who completed the task. I then averaged each difference to determine how much each speaker deviated from Birdsong's and Donaldson's NSs. This analysis also serves to determine how closely my four NSs (including the Pau bilinguals) match with the 20 NSs from the Birdsong and Donaldson studies combined. I then subtracted the best theoretically possible score from each speaker's overall average (since items were judged as integers on a 1-5 scale and thus could not be identical to non-integer averages). In this way, a speaker who theoretically assigned the same rating as the average of these NSs for each item (rounded up or down depending on the non-integer average) would then have an average deviation of 0.0, while a speaker who consistently rated each item one point higher (or lower) than the NS average would have an average deviation of 1.0. The smaller the deviation, the more closely the speaker matches with NS intuitions on complex grammar structures from the Birdsong and Donaldson studies.

As a group, the Near-NSs in Pau averaged 0.789 deviation, which matches very closely with the near-native average deviation in Donaldson (0.753); a two-tailed *t*-test determined a non-

significant difference between these two groups (p = .63). Near-NSs in Lille averaged 0.862 deviation, a larger difference compared with Donaldson's group but nevertheless not significant (p = .09). These results again suggest that the speaker populations for each site closely fit the profile of near-native speakers established by Donaldson (contra the pilot study participant group; cf. Appendix A.3). A *t*-test also determined a non-significant difference between my two Near-NS groups (p = .39). For individual speaker scores, the results are organized by site in Table 3-8.

Table 3-8. Average deviation from Birdsong and Donaldson NSs on AJT

Speaker ID: Pau	**Avg. deviation from NSs**	**Speaker ID: Lille**	**Avg. deviation from NSs**
1P	0.711	1L	0.743
2P	0.383	2L	0.852
3P	0.706	3L	0.885
4P	1.077	6L	0.847
5P	0.760	7L	0.845
6P	0.872	8L	0.746
8P	0.849	9L	1.119
9P	0.851	SaE	0.868
10P	0.889	JeE	0.684
Ch	0.706	CaF	0.422
Fr	0.608	KeF	0.619

Note, first, that three of the four NSs in the current study who completed the AJT (CaF, KeF, Fr) have the lowest average deviation other than speaker 2P, with bilingual Ch trailing these speakers in addition to near-native JeE. Not only would these low deviations for the NSs be expected, but the lowest deviation for speaker 2P serves to corroborate my informal observation that this speaker appeared to be the most nativelike of all the Near-NSs recruited.

3.5.2.7 Results: Language security index

As with the *c*-test for SA learners, whereas the AJT may be a good measure of Near-NSs' grammatical ability in French, the language security index can also capture other factors involved

in a speaker's overall competence and motivation for speaking and learning the language, which may also influence the speaker's sociolinguistic awareness. The questions in Donaldson's (2008) original questionnaire concerning passing as a native speaker and attempting to pass as a native speaker were based on work by Piller (2002), who found that some near-natives may choose to preserve certain features of their L2 speech distinguishing them from the target community (see also Gnevsheva (2017) and Forsberg Lundell et al. (2014)). Such motivations can certainly play a part in the development of a learner's overall speaking proficiency. Table 3-9 provides the language security index for Near-NSs based on results from their background questionnaires.

Table 3-9. Measure of language security in Near-NSs

Speaker ID	**Abilities in French**					**Motivations in French**		**Security Index (1-5)**
	Read	Hear	Write	Speak	Accent rating	Pass as NS	Attempt to pass as NS	
1P	4	4	3	4	4	3	3	3.6
2P	5	5	5	5	5	5	1	4.4
3P	4	4	4	4	3	1	1	3.0
4P	5	5	4.5	5	4.5	3.5	2	4.2
5P	5	4.5	4	5	4	3	1	3.8
6P	5	5	5	5	4	3	5	4.6
7P	5	5	5	5	5	5	3	4.7
8P	4	4	4	5	4	4	5	4.3
9P	4	4	3	4	3	2	1	3.0
10P	5	5	5	5	4	3	1	4.0
1L	4.5	4	4	4	3	2	2	3.4
2L	4	4	4	4	4	1	1	3.1
3L	4	4	3.5	4	3	2	3	3.4
4L	5	5	5	5	4.5	3	5	4.6
5L	4	4	4	4	4	3	3	3.7
6L	3	4	2	4	3	2	1	2.7
7L	5	5	4	5	4	1	1	3.6
8L	5	5	4	4	4	3	3	4.0
9L	5	5	4	5	4	2	3	4.0
Average	4.5	4.4	4.1	4.5	4.0	3.1	2.3	3.8

As expected, Near-NSs report higher language security overall compared with SA learners (3.8 versus 3.3), though it is possible that SA learners may have "inflated" their self-reported abilities and Near-NSs may have underestimated their abilities by comparison. Nevertheless, certain scores on the language security index may more accurately reflect learner behavior concerning sociolinguistic variation compared with results on the AJT, given each speaker's motivations for using the language in their native speaker community. With such caveats in mind, this measure will be used along with the AJT in order to find potential correlations with the use of the sociolinguistic variables to be analyzed in the current study.

3.6 Corpus

As for oral production data, over all sites and speaker groups, the corpus consists of 67 different recorded conversations. Eight of these conversations consisted of participants who were excluded from the analysis due to not meeting proficiency criteria. The remaining 59 conversations include the 10 Near-NSs retained in Pau (20 conversations: 10 English-identity-NS/Near-NS dyads, 10 French-identity-NS/Near-NS dyads), the 9 Near-NSs retained in Lille (18 conversations: 9 NS/Near-NS dyads, 9 near-native/Near-NS dyads), and the 8 SA learners (21 conversations: 8 SA/NS dyads, 8 SA/near-native dyads, 5 SA/SA dyads[15]). The approximate average length of conversation for each type of dyad is provided in Table 3-10.

[15] The SA-SA conversations were originally conceived as consisting of four conversations between four pairs of learners: 1S-2S, 3S-4S, 5S-6S, 7S-8S. Due to scheduling limitations, however, the learners could not be paired in this way, and an extra conversation was necessary: 1S-2S, 3S-4S, 3S-5S, 6S-7S, 7S-8S. This resulted in two learners (3S and 7S) each recording a second SA-SA conversation.

Table 3-10. Average conversation length of each dyad type

Dyad type	No. of conversations	Average conversation length
Near-NS/French identity bilingual (Pau)	10	35 min, 31 sec
Near-NS/English identity bilingual (Pau)	10	36 min, 26 sec
Near-NS/NS (Lille)	9	34 min, 59 sec
Near-NS/near-native (Lille)	9	33 min, 47 sec
SA learner/NS	8	22 min, 42 sec
SA learner/near-native	8	21 min, 20 sec
SA learner/SA learner	5	19 min, 26 sec

These 59 conversations, representing nearly 30 hours of recorded speech, were subsequently transcribed for data analysis. Aware of Coveney's estimation (2002: 22) that five minutes of recorded speech requires at least one hour of manual transcription time, I sought automatic transcription tools in order to facilitate this process. The only application that was reasonably successful was VoiceNote II, a Chrome browser extension that allows for automatic transcription of many languages, including French. With this extension running within the browser window, and playing the audio file with the appropriate playback settings, automatic transcription can be obtained with a limited degree of accuracy, which allowed for a rough transcription of some of the recordings. However, SA learner speech was transcribed much less accurately than Near-NS or NS speech, certainly due to non-nativelike syntax and phonology. Since these learners' speech pace was relatively slow, I decided to transcribe their audio recordings manually. As for Near-NSs, because each participant had their own microphone and audio recorder, VoiceNote II often did not pick up the audio from the interlocutor, so a second run with the interlocutor's audio file had to be conducted. Despite these issues, for certain recordings this software was able to render large portions of the audio stream into text. Afterward, I cleaned up the output and fixed inaccuracies manually, using the VLC software program for listening to the recordings, as

Donaldson (2017) had done. Near the end of the transcription process, I became aware of other software that streamlined the task somewhat. Express Scribe is a program that has easy-to-use tools for altering the playback speed, hotkeys for pausing and rewinding the audio, and, importantly, the ability to use these hotkeys with the program running in the background (unlike with VLC), allowing a word processor program window to remain in focus for continuous typing.

The resulting transcribed corpus consists of over 278,000 words. For the purposes of this study, I did not need a finely detailed transcription indicating all pauses or indicating which speech was overlapping. I excluded most backchannel feedback from the interlocutor (e.g., *mm-hmm, oui, ouais*). This otherwise complete transcription of the entire corpus facilitated the following tasks: 1) Analysis of quantitative dominance across all conversations, as detailed in the following section; 2) Verification of the existence of other informal features, as detailed beginning in section 3.6.2; and 3) Cross-checking of items in the factor groups for the subsequent variationist analyses, as detailed in Chapters 4 and 5.

3.6.1 Quantitative dominance

With dyadic conversations of the type collected in this corpus, where speakers of different proficiency levels interact in each dyad, it is important to determine to what degree each speaker participated in each conversation. A measure of quantitative dominance (Itakura, 2001) was obtained by counting the words each speaker produced. Glahn (1993) finds that in dyads composed of a native speaker and an advanced non-native speaker, there tends to be minimal native speaker dominance. Therefore, I examined the quantitative dominance in all dyads of native speakers (including bilinguals in Pau) and Near-NSs in my corpus; for sake of comparison, I also examined the quantitative dominance in Near-NS and near-native interlocutor conversations, as well as all conversations with the SA learner group. For all analyses, I followed Donaldson's (2008)

procedure by excluding from my original transcriptions all external noise, nonlexical hesitations, backchannels, pauses, and indecipherable passages.

3.6.1.1 Quantitative dominance: Near-NS conversations

Starting with the Near-NS conversations, Table 3-11 reports the quantitative dominance for the dyads in Pau.

Table 3-11. Quantitative dominance in Near-NS/NS dyads in Pau

		Near-NSs				**Native interlocutors**		
Dyad	Total words	Speaker ID	No. of words	% of pro-duction		Speaker ID	No. of words	% of pro-duction
1	5574	1P	2816	50.5	=	ThF	2758	49.5
2	5903	2P	3321	56.3	=	ChF	2582	43.7
3	6683	3P	2821	42.2	=	FrF	3862	57.8
4	6132	4P	2696	44.0	=	ChF	3436	56.0
5	6939	5P	4550	65.6	>	FrF	2389	34.4
6	4930	6P	1958	39.7	=	ChF	2972	60.3
7	6514	7P	4168	64.0	>	FrF	2346	36.0
8	6316	8P	3951	62.6	=	FrF	2365	37.4
9	6114	9P	2128	34.8	<	ChF	3986	65.2
10	6736	10P	2075	30.8	<	ChF	4661	69.2
11	5886	1P	2833	48.1	=	ChE	3053	51.9
12	5378	2P	3400	63.2	=	FrE	1978	36.8
13	6631	3P	2294	34.6	<	ThE	4337	65.4
14	5404	4P	3984	73.7	>	ThE	1420	26.3
15	6518	5P	2881	44.2	=	ChE	3637	55.8
16	6089	6P	2283	37.5	=	FrE	3806	62.5
17	6738	7P	3135	46.5	=	ChE	3603	53.5
18	7115	8P	1441	20.3	<	ChE	5674	79.7
19	5924	9P	2900	49.0	=	FrE	3024	51.0
20	6619	10P	3831	57.9	=	FrE	2788	42.1
Total	124,143		59,466	47.9	=		64,677	52.1

As this table shows, there is wide variation in terms of which kind of speaker (native or non-native) dominates in a particular conversation; however, overall, there is minimal native speaker

dominance (52.1% of all words), and a paired samples *t*-test comparing the percentages of words produced for each group revealed that this difference between native speakers and Near-NSs is not significant ($t(38) = -0.90$, $p = .374$, $SD = 13.58$). This difference may be rendered even less substantial given that NSs were generally observed to speak more quickly than Near-NSs overall, increasing their word count relative to the amount of time spent speaking. Following Donaldson (2008), I also determined whether there was dominance of one speaker in any of the dyads (indicated by the symbols '<' and '>') when the speaker's percentage of word production was above one standard deviation ($SD = 13.58$). Speakers with over 63.58% of word production in each dyad were thus considered dominant, and dyads in which neither speaker produced over 63.58% were considered evenly matched for dominance (indicated by the symbol '='). In three dyads (5, 7, 14), the Near-NS was dominant; in four dyads (9, 10, 13, 18), the NS was dominant. In none of the dyads was the same Near-NS dominated by both NS interlocutors. The seven dyads (out of 20) showing dominance of one speaker are comparable to the number obtained by Donaldson (where three of his 10 dyads showed dominance of one speaker). Furthermore, there was minimal difference in the NS interlocutors when adopting an English identity (53.5% of all words) compared with a French identity (50.7% of all words).

Concerning the dyads in Lille, Table 3-12 reports the results for the conversations between Near-NSs and native speakers. There is comparatively slightly more dominance by NSs in these dyads (55.9% of all words), though as in Pau, the faster speaking pace (in terms of words per minute) of these NSs likely accounts for part of this difference. A paired samples *t*-test revealed that the difference in percentage of word production between these two groups is not significant ($t(16) = 1.93$, $p = .089$, $SD = 9.39$), though much closer to significance than the Near-NS/NS dyads in Pau.

Table 3-12. Quantitative dominance in Near-NS/NS dyads in Lille

		Near-NSs				Native interlocutors		
Dyad	Total words	Speaker ID	No. of words	% of production		Speaker ID	No. of words	% of production
1	7190	1L	2931	40.8	=	CaF	4259	59.2
2	6898	2L	2329	33.8	<	CaF	4569	66.2
3	7090	3L	4036	56.9	=	CaF	3054	43.1
4	6959	4L	3333	47.9	=	CaF	3626	52.1
5	4015	5L	1333	33.2	<	CaF	2682	66.8
6	5527	6L	3211	58.1	=	CaF	2316	41.9
7	6148	7L	2877	46.8	=	CaF	3271	53.2
8	6340	8L	2223	35.1	<	KeF	4117	64.9
9	5630	9L	2331	41.4	=	KeF	3299	58.6
Total	55,797		24,604	44.1	=		31,193	55.9

Using the same type of measurement for quantitative dominance as in the Pau dyads (where dominance is 50% of words plus one standard deviation, or 59.39%), of these nine conversations, three were dominated by NSs (2, 5, 8).

Broadly speaking, the native speakers, whether monolinguals in Lille or bilinguals in Pau, show minimal dominance over Near-NSs, as has been found in previous studies (Donaldson, 2008; Glahn, 1993). The range of quantitative dominance by Near-NSs (low of 20.3% and high of 73.7%) is also comparable to the range found in Donaldson (low of 21.5% and high of 74.1%). These dyads can thus be considered representative of the type of dominance one would expect in such groupings of speakers of different L1s.

Focusing specifically on the Near-NS/near-native dyads, which includes nine conversations in Lille, there is no quantitative dominance in favor of one group over the other, with both kinds of speakers each producing essentially 50% of all words (see Table 3-13).

Table 3-13. Quantitative dominance in Near-NS/near-native dyads in Lille

		Near-NSs				Near-native interlocutors		
Dyad	Total words	Speaker ID	No. of words	% of production		Speaker ID	No. of words	% of production
1	4112	1L	1706	41.5	=	SaE	2406	58.5
2	4387	2L	1955	44.6	=	SaE	2432	55.4
3	4224	3L	2154	51.0	=	SaE	2070	49.0
4	4889	4L	2803	57.3	=	JeE	2086	42.7
5	6131	5L	2033	33.2	<	SaE	4098	66.8
6	4641	6L	2557	55.1	=	JeE	2084	44.9
7	4694	7L	3158	67.3	>	JeE	1536	32.7
8	4680	8L	2242	47.9	=	SaE	2438	52.1
9	4327	9L	2434	56.3	=	SaE	1893	43.7
Total	42,085		21,042	50.0	=		21,043	50.0

As expected, the paired samples *t*-test comparing word production across both speaker groups shows a highly insignificant difference between the two groups ($t(16) = .0003$, $p = .99975$, $SD = 10.03$). Two of the dyads showed quantitative dominance of one speaker (60.03% of all words): the Near-NS in dyad 5 and the near-native interlocutor in dyad 7.

3.6.1.2 Quantitative dominance: SA learner conversations

The dyads involving SA learners were also analyzed for quantitative dominance. One might expect more dominance by either the near-native or native interlocutor (or both), given their much higher proficiency compared with SA learners; however, the differences are quite minimal. Table 3-14 provides the SA learners' production with the NS interlocutor.

Table 3-14. Quantitative dominance in SA learner/NS dyads

		SA learners				Native interlocutor (SoF)	
Dyad	Total words	Speaker ID	No. of words	% of production		No. of words	% of production
1	2075	1S	1076	51.9	=	999	48.1
2	2515	2S	1478	58.8	>	1037	41.2
3	2495	3S	1377	55.2	>	1118	44.8
4	3755	4S	1804	48.0	=	1951	52.0
5	2438	5S	1239	50.8	=	1199	49.2
6	2472	6S	1072	43.4	<	1400	56.6
7	2403	7S	1145	47.6	>	1258	52.4
8	2263	8S	1118	49.4	>	1145	50.6
Total	20,416		10,309	50.5	=	10,107	49.5

A *t*-test comparing word production across both kinds of speakers shows a highly insignificant difference ($t(14) = -0.296$, $p = .776$, $SD = 4.77$). Given the small standard deviation of these percentages, a majority of these conversations show quantitative dominance of more than 54.77% of total words (dyads 2, 3, 6, 7, 8); however, the fact that none of the dyads show substantial dominance suggests that these conversations allowed for roughly equal participation by both speakers. (Contrast with Porter (1986), whose NS interlocutors averaged 62% of the total words in conversation with learners.) The fact that four of these conversations show slight dominance by the SA learner also indicates that these learners were quite capable of sustaining extended conversation.

Table 3-15 summarizes the SA learners' production with the near-native interlocutor (AmE).

Table 3-15. Quantitative dominance in SA learner/near-native dyads

		SA learners				Near-native interloc. (AmE)	
Dyad	Total words	Speaker ID	No. of words	% of production		No. of words	% of production
1	2569	1S	1385	53.9	=	1184	46.1
2	1998	2S	1099	55.0	=	899	45.0
3	3123	3S	1883	60.3	>	1240	39.7
4	2674	4S	1157	43.3	=	1517	56.7
5	3291	5S	1387	42.1	=	1904	57.9
6	2514	6S	954	37.9	<	1560	62.1
7	2710	7S	1118	41.3	=	1592	58.7
8	2333	8S	1412	60.5	>	921	39.5
Total	21,212		10,395	49.0	=	10,817	51.0

A *t*-test comparing word production across both kinds of speakers reveals a highly insignificant difference ($t(14) = 0.303$, $p = .771$, $SD = 9.12$). There are larger variances in these dyads compared to the SA learner/NS dyads, even though the overall percentages between SA learners and the near-native interlocutor are nearly identical. Speakers 3S and 8S show quantitative dominance (more than 59.12% of all words) over AmE, while AmE shows dominance over only one learner (6S).

The native and near-native interlocutors thus appear to have fulfilled their roles as conversational partners. Broadly, neither interlocutor dominated the conversations, providing an environment favorable for learners at this proficiency level to express themselves in extended discourse. Likewise, these interlocutors did not simply act as an interviewer; even though they tended to ask the majority of questions, they also initiated reflections on their own interests and experiences concerning the topics discussed.

Finally, the quantitative dominance in the five conversations conducted by the SA learner/SA learner dyads is summarized in Table 3-16. Since each dyad consists of learners in the same speaker group, between-group comparisons are not possible. The place of the learners in this

table (appearing on the left or right side of the table) are simply determined by the number of the Speaker ID, and the dyads are listed in the order in which the conversations were recorded. We can, however, determine the standard deviation of the percentage difference in word production by these speakers (*SD* = 7.66). Based on this calculation, only one dyad (5) produced quantitative dominance by one speaker (8S) with over 57.66% of the total words. Note also that, due to the shorter average length of these conversations, and due to the slower speaking pace of the SA learners, the number of words in each conversation is much lower than in SA learner dyads involving speakers SoF and AmE.

Table 3-16. Quantitative dominance in SA learner/SA learner dyads

		SA learner				**SA learner**		
Dyad	Total words	Speaker ID	No. of words	% of production		Speaker ID	No. of words	% of production
1	1476	1S	671	45.5	=	2S	805	54.5
2	2287	3S	1269	55.5	=	4S	1018	44.5
3	2539	3S	1121	44.2	=	5S	1418	55.8
4	2158	6S	1211	56.1	=	7S	947	43.9
5	2018	7S	793	39.3	<	8S	1225	60.7
Total	10,478							

The quantitative dominance in these learner dyads reveals little overall difference compared with the dominance in other dyads with SA learners, and little difference (and smaller ranges) in terms of quantitative dominance compared with dyads involving Near-NSs. Broadly speaking, then, the structure of these conversations is comparable across the different speaker groups, across the sampling sites, and can be compared to other studies (e.g., Donaldson, 2008) with a similar type of oral production task.

3.6.2 Informal features: Near-NSs

What cannot be automatically assumed, however, is whether the speakers in each dyad adopted an informal style for these conversations. As in Donaldson (2008), identification of certain features characteristic of informal French that are uttered by each speaker can determine how informally, linguistically speaking, these speakers treated the conversation tasks. Given the type of interactions in this corpus, it would furthermore be instructive to examine each speaker's use of these features across each conversation (recall Table 3-1 in section 3.1), in order to verify which informal features appear in both conversations, and to ensure that the speakers did not treat one conversation as substantially more formal overall due to differences in the interlocutors. According to Donaldson's criteria, if the speaker produced the feature at least once during the conversation, it was considered as present in that speaker's speech. If an informal feature is attested, we can then later examine in more detail the *frequency* in which this feature appears in variable contexts, and evaluate to what extent characteristics of the interlocutors (as well as other social, and linguistic, factors) may condition the frequency of the informal variant.

I therefore begin with a series of tables modeled on Donaldson's tables (2008: 125-126) for Near-NSs and their interlocutors, where a plus sign indicates the appearance of a feature (appearing at least once), and the word "no" indicates that the feature was not used by the speaker. Table 3-17 and Table 3-18 show the features used by the Near-NSs in Pau, according to interlocutor type (note that only the 10 Near-NS speakers retained for analysis are included here). Table 3-19 and Table 3-20 provide the distribution by the nine Near-NSs in Lille. Finally, Table 3-21 indicates the use of these features by the interlocutors for the Near-NSs, over all sites; for the bilinguals (Ch, Fr, Th), tokens are broken down according to whether they were uttered under their English (E) or French (F) identity; for the remaining interlocutors, recall that SaE and JeE are near-

native (L1 English) speakers, while CaF and KeF are native French speakers. Note also that an asterisk next to the speaker's ID indicates that the speaker used *vous* to address her interlocutor.

Table 3-17. Informal features of Near-NSs in Pau (with English-identity interlocutor)

Feature	**1P***	**2P***	**3P***	**4P***	**5P**	**6P***	**7P**	**8P**	**9P**	**10P**
Ne-deletion	+	+	+	+	+	+	+	+	+	+
Interrogatives	+	+	+	+	+	+[16]	+	+	+	+
Truncation	+	+	+	no	+	no	+	no	+	+
Pronoun reduction	(+)	+	+	+	+	+	+	+	+	+
/l/-deletion	+	+	+	+	+	+	+	+	+	+
Object drop	+	no	+	+	+	+	+	+	+	+
Pragmatic particles	+	+	+	no	+	no	+	no	+	no
On for *nous*	+	+	+	+	+	+	+	+	+	+
Vocabulary	+	+	+	+	+	+	+	+	+	+

Table 3-18. Informal features of Near-NSs in Pau (with French-identity interlocutor)

Feature	**1P***	**2P***	**3P***	**4P***	**5P**	**6P***	**7P**	**8P**	**9P**	**10P**
Ne-deletion	+	+	+	+	+	+	+	+	+	+
Interrogatives	+[17]	+	+	+	+	+	+	+	+	+
Truncation	no	+	no	+	no	no	+	no	+	+
Pronoun reduction	+	(+)	+	+	(+)	+	+	+	+	+
/l/-deletion	+	+	+	+	+	+	+	+	+	+
Object drop	no	+	+	+	no	no	+	+	no	+
Pragmatic particles	+	+	+	no	+	+	+	+	+	no
On for *nous*	+	+	+	+	+	+	+	+	+	+
Vocabulary	+	+	+	+	+	+	+	+	+	+

[16] This speaker produced one formal interrogative with subject-verb inversion.

[17] This speaker used almost exclusively ESQ (*est-ce que*) constructions in her questions but no inversions.

Table 3-19. Informal features of Near-NSs in Lille (with near-native interlocutor)

Feature	1L	2L	3L	4L	5L	6L	7L	8L	9L
Ne-deletion	+	+	+	+	+	+	+	+	+
Interrogatives	+	+	+	+	+	+	+	+	+
Truncation	+	no	+	+	no	+	no	+	+
Pronoun reduction	+	+	+	+	+	+	+	+	+
/l/-deletion	+	+	+	+	+	+	+	+	+
Object drop	+	no	+	no	no	no	no	+	+
Pragmatic particles	no	+	no	+	+	no	+	+	+
On for *nous*	+	+	+	+	+	+	+	+	+
Vocabulary	+	+	+	+	+	+	+	+	+

Table 3-20. Informal features of Near-NSs in Lille (with NS interlocutor)

Feature	1L	2L	3L	4L	5L	6L	7L	8L	9L
Ne-deletion	+	+	+	+	+	+	+	+	+
Interrogatives	+	+	+	+	+	+	+	+	+
Truncation	+	no	+	+	no	+	no	no	no
Pronoun reduction	+	+	+	+	+	(+)	+	+	+
/l/-deletion	+	+	+	+	+	+	+	+	+
Object drop	+	+	+	+	+	+	no	+	+
Pragmatic particles	no	+	no	+	+	no	+	+	+
On for *nous*	+	+	+	+	+	+	+	+	+
Vocabulary	+	+	+	+	+	+	+	+	+

Table 3-21. Informal features of interlocutors for Near-NSs (both sites)

Feature	ChE	ChF	FrE	FrF	ThE*	ThF*	SaE	JeE	CaF	KeF
Ne-deletion	+	+	+	+	+	+	+	+	+	+
Interrogatives	+	+	+	+	+	+	+	+	+	+
Truncation	+	+	+	+	+	no	+	no	+	no
Pronoun reduction	+	+	+	+	+	+	+	+	+	+
/l/-deletion	+	+	+	+	+	+	+	+	+	+
Object drop	+	+	+	+	+	no	+	+	+	+
Pragmatic particles	+	+	+	+	+	+	+	+	+	+
On for *nous*	+	+	+	+	+	+	+	+	+	+
Vocabulary	+	+	+	+	+	no	+	+	+	+

Note: Ch, Fr, Th = bilinguals; SaE + JeE = near-native speakers; CaF + KeF = native speakers.

Compared with Donaldson's tables, the major modification of feature distribution in my tables concerns what are called "pragmatic particles," following Beeching's (2001) terminology for *enfin*, though they are often identified more generally as "discourse markers." Donaldson included *enfin* (Beeching, 2001, 2011) and *hein* (mentioned briefly in Valdman, 1982), which may constitute part of what, in his study on right-dislocation (Donaldson, 2011b), he admits are a somewhat random selection of informal features (though he maintains his argument that such features are among the most discussed aspects of spoken French). As introduced in section 3.1, there are other discourse particles not included here that appear in spontaneous, informal, Hexagonal French, such as specific uses of *quoi, genre, machin,* and *bref*, all of which I included in my identification of pragmatic particles. This is not an exhaustive list; see, e.g., Haileselassie (2015) for *voilà* and Reaves (2020) for an extensive overview. However, since the relevance for identifying these pragmatic particles is to determine whether speakers incorporate at least one kind of this particular feature, the distribution of use of specific pragmatic particles in each participant is of secondary importance, and I therefore collapse all such tokens into one category for the tables in this section. I refer the reader to Appendix B for a more detailed discussion on the background and use of each of these pragmatic particles for all speaker groups.

I conclude this section with some general observations on the use of pragmatic particles as they relate to the current study. Recall that, as Donaldson (2011b) mentions, formality exists on a continuum, containing a range of formal and informal features. The appearance of several different pragmatic particles would likely indicate a more informal style, but the absence of a specific particle may not necessarily indicate a less formal style. As the distribution in Appendix B indicates, not all native speakers use all of the pragmatic particles in their informal speech, though usage of any such particles could be cited as evidence of a speaker leaning toward a more informal

style. Moreover, there does not appear to be a great deal of unequal distribution of these particles in Near-NSs according to interlocutor type: generally, when a speaker uses a particular particle in one conversation, she tends to use it in the other. As for interlocutors, whose totals encompass multiple conversations with the same type of speaker, there was no obvious evidence of unusual distribution or clustering of these particles with one Near-NS or a particular group of Near-NSs.

More broadly, these particles are distributed over what may be considered a continuum of formality. Using *vous* to address one's interlocutor does not necessarily inhibit the use of pragmatic particles, as speaker 2P used *vous* but had the largest range of pragmatic particles of all Near-NSs. On the other hand, the absence of these pragmatic particles (as in speakers 1L, 3L, and 6L) does not necessarily indicate a more formal conversation, as evidenced by, for example, 1L's low *ne*-retention rate, or 6L's extensive use of informal vocabulary. Near-NSs generally have a more limited range of pragmatic particles compared with NSs, though this distribution may not be noticeable by interlocutors or may not necessarily be an indicator of non-native speech. It seems that some learners "latch on" to a particular particle once they are comfortable using it in certain settings (whether this is conscious or unconscious on their part, as may be the case for native speakers), and usually only after a period of extensive contact with native speakers (see Reaves (2020: 81) for an overview of studies on learners overusing specific discourse markers).

3.6.2.1 Additional analysis of informal features in Near-NSs

Returning to the distribution of the remaining informal features, with the caveat that the "Pragmatic particles" category encompasses multiple particles in my tables, there are two main differences between my Near-NSs and Donaldson's speakers. First, fewer of my speakers use truncation and object drop (or null objects), though in Donaldson's study, these were the only two features that were missing from some of the speakers (truncation for three of the NSs; null objects

for two of the near-natives and three of the NSs). These tokens may be considered more context-dependent than other features; as discussed in section 3.1, it is possible to have an entire conversation in an informal style and not produce these features, depending on the topic of conversation. Furthermore, for the Near-NSs, the presence/absence of these tokens in each table is based on between 30-40 minutes of conversation, compared with 45-58 minutes for Donaldson's speakers; combining both conversations for each Near-NS results in the appearance of nearly all informal markers for each of these speakers.

Second, some of my speakers used *vous* as the pronoun of address, whereas Donaldson's speakers, interacting with a spouse or a close friend, used *tu*. The choice of *vous* as the pronoun of address by some of my speakers does not appear to meaningfully alter the distribution of informal features in these conversations. None of the Near-NSs lacked more than two informal features in any of their conversations, regardless of the chosen pronoun. Speakers using *vous* lacked one informal feature on average, essentially the same average as Near-NSs in Lille, none of whom used *vous*. The single exception may be Th's lack of three different informal markers in his lone conversation under his French identity, which may be more due to the large age difference with his interlocutor rather than the choice of address pronoun; using *vous* under his English identity did not inhibit production of any informal markers.

Another somewhat minor difference concerns pronoun reduction. For this category, Donaldson (2008) references George (1993), whose study focuses largely on the informal lexicon and truncation but briefly mentions other aspects of informal French, including phonetic elision with two subject pronouns: 2SG *tu* (e.g., *t'as* for *tu as*, 'you have') and 3SG.MASC *i* for *il* ('he'). Since /l/-deletion with pronoun *il* constitutes a separate category in Donaldson's table of informal features (cf. Table 3-1), Donaldson (personal communication, July 16, 2019) indicates that most

of his tokens in the category of pronoun reduction concerned *tu* (such as *t'as* for *tu as* or *t'es* for *tu es*); however, other examples of pronoun reduction may include elision of the entire impersonal pronoun *il* in *(il) faut* ('one must')[18] as well as reduced forms of *vous*, where the vowel /u/ is elided.[19] All but two of my Near-NSs who used *tu* with their interlocutors showed reduction of *tu* (speakers 1L and 6L). Several others elided *il* with forms of *falloir* (including speaker 1L), and two other speakers (ThF and 3P) had clear examples of elided vowels in *vous*.

After accounting for these forms, there remained several Near-NSs who did not produce pronoun reduction as described above. Though not mentioned in previously cited literature, one could also include reduced forms of *je* in this category. If the scope of pronoun reduction is extended to, for example, vowel elision and devoicing of *je*, e.g. /ʃɥi/ (sometimes transcribed as *chui*) for standard *je suis* (/ʒəsɥi/), then pronoun reduction can be attested in all other Near-NSs. Such speakers whose only pronoun reduction was in *je* are noted by the (+) symbol in Table 3-17 through Table 3-21; this was observed for speakers 1P, 2P, 5P, and 6L.

With these differences in mind, two general observations can be made in comparison with Donaldson's near-native speakers. Either my Near-NSs have adopted, broadly speaking, a somewhat more formal style in these conversations than Donaldson's (but still falling more on the informal end of the continuum, based on the distribution of informal features), or my two groups are, broadly speaking, somewhat less nativelike than Donaldson's speakers, failing to adopt certain informal features due to lower proficiency. There may, in fact, be a certain component of both factors. Certainly, the fact that Donaldson's near-natives chose their own (native) interlocutors (and the resulting informality implied by these dyads) may address the discrepancy in distribution.

[18] In addition to *il faut*, elision of *il* may occur with other forms of *falloir* (e.g., *(il) faudrait*, 'one should').
[19] In *liaison* contexts, both the vowel and the onset consonant /v/ may be elided, leaving only the /z/ of *liaison* (e.g., *Z'avez vu ?* ('Have you seen?') for *Vous avez vu ?*.

Though measures of syntactic competence as determined by an Acceptability Judgment Task may not necessarily correlate with more nativelike use of informal features, the AJT results as reported in section 3.5.2.6 seem to indicate that my Near-NSs are comparable in terms of proficiency. These differences aside, there is sufficient evidence in the distribution of informal features that these Near-NSs largely treated the conversation tasks as informal to the extent that my methodology allowed—and, for some speakers, to the extent that their pronouns address allowed.

3.6.3 Informal features: SA learners

Though an analysis of the aforementioned informal features was likely not intended by Donaldson for learners at the proficiency level of my SA learners, it can nevertheless be instructive to analyze these learners' speech to see which features are present at this level, in addition to quantifying in some way the level of formality adopted by the native and near-native interlocutors for these SA learners. Table 3-22 shows, for the participants in the SA learner conversations, the distribution of the same markers as those analyzed in the previous section with Near-NSs.

Table 3-22. Informal features of SA learners and interlocutors

Feature	**1S**	**2S**	**3S**	**4S**	**5S**	**6S**	**7S**	**8S**	**AmE**	**SoF**
Ne-deletion	+	+	+	+	+	(no)	(no)	+	+	+
Interrogatives	+	+	+	+	+	+	+	+	+	+
Truncation	no	(no)	+	+	(no)	+	+	(no)	+	+
Pronoun reduction	no	no	no	no	no	no	no	no	+	+
/l/-deletion	no	(no)[20]	no	+	no	no	+	+	+	+
Object drop	no	+	no	+	no	+	+	no	+	+
Pragmatic particles	no	no	no	no	no	no	no	no	+	+
On for *nous*	no	no	+	no	+	no	+	+	+	+
Vocabulary	+	+	+	+	+	no	+	+	+	+

[20] This speaker produced one utterance of *il y a* where /l/-retention could not be confirmed; in addition, the speaker produced /l/-deletion in the lexicalized expression *s'il vous plait* ('please') on several occasions.

As may be expected, SA learners categorically lack several of these informal features, whereas nearly all of them use *ne*-deletion, informal interrogatives, and informal vocabulary to some extent, suggesting at least some competence at producing a variety of informal features. As will be discussed in more detail in Chapter 4, the "(no)" between parentheses indicated for speakers 6S and 7S in the *ne*-deletion row is due to the fact that these two speakers each produced only one token of *ne*-deletion overall, and these two tokens are marginal (for speaker 6S, there was a slight pause between the verb and *pas*, while for speaker 7S, the only deletion was an immediate and identical replication of his interlocutor's utterance *c'est pas cher*, 'it's not expensive'). As for interrogatives, all learners use the informal *wh* in-situ (e.g., *Tu pars quand?* 'When are you leaving?'), though they tend to use the more neutral *est-ce que* (e.g., *Quand est-ce que tu pars?*) more often than near-native and native speakers do. Excluding fixed expressions, four learners (2S, 3S, 5S, 8S) produced formal interrogative inversion (e.g., *Quand pars-tu?*), though only one learner (2S) produced more than one of these tokens ($n = 3$). In general, however, learners did not ask many questions directed to the near-native and native interlocutors. Regarding truncation, the "(no)" between parentheses indicates that the only truncation by these learners was the word *sympa* ('nice'), which is an abbreviated form of *sympathique.* However, it is common for learners of French to learn the abbreviated form *sympa* as a lexical item early in classroom instruction, so it is not clear if they distinguish *sympa* as being more colloquial than *sympathique* (which was produced by none of the learners). Otherwise, the only other truncations in SA learners included the word *resto* ('restaurant'), produced by 3S, 4S, and 7S, and *muscul* (by 6S) in the expression *faire de la muscul* ('to work out / to do weightlifting'). Concerning *on/nous*, all learners used subject *nous* at least once, and while all of them used informal *on* at least once, only four used it in a context where *nous* could have clearly been substituted. It is not always clear, however,

whether *on* is being used exclusive or inclusive of the speaker (where variation with *nous* is possible only in the latter case). True examples of object drop (or null objects) can also be difficult to determine in learners at this proficiency level; production of a verb without an object complement may occur where a speaker did not intend to drop an NP complement (often due to lexical retrieval delays), or where a speaker drops an intended object clitic due to processing difficulties in syntactic computations or feature agreement. Nevertheless, several SA learners produced what were clearly instances of felicitous object drop.

As for the interlocutors, a review of the conversations indicated that both the native and near-native speaker adopted an informal conversational tone, as mentioned in section 3.5.1.2. However, the appearance of all of these informal features in the above table gives quantitative support to this observation. Note that the two pragmatic particles identified for both interlocutors were the most commonly produced particles in the Near-NS groups: *enfin* and *quoi* (see Appendix B). AmE also produced much more informal vocabulary than SoF, who may have been hesitant to use such vocabulary out of concern of lack of comprehension (though SA learners asked for more clarification of lexical items in conversation with SoF compared with AmE).

Overall, it is clear that these conversations created an environment for the use of multiple kinds of informal features in SA learners as well as their interlocutors, suggesting the adoption of an informal style for both speakers in these conversations. The absence of other informal features in learners may simply be due to a lack of integration of these features into their active production (despite use of all of these features by their more proficient interlocutors, and despite presumable frequent input of such features from other interactions with native speakers), rather than a conscious choice to exclude such items from their speech. Moreover, I make no presuppositions that these learners would be capable of accurately or consistently producing the formal variants of

relevant features if an oral production task requiring adoption of a formal style were presented. What this analysis does suggest is that these learners are capable of style-switching from the type of language to which they are exposed in typical classroom environments early in their study of the language, toward a style that they perceive to be appropriate given my task instructions and given the environment in which they were recorded.

3.6.4 Which sociolinguistic variables to analyze?

Having established that the corpus obtained for the current study contains recordings evidenced to be of an informal nature, we can move beyond basic detection of whether certain stylistically conditioned variables appear in their informal variants to an analysis of the frequency and the distribution of these variants in all contexts in which they have the potential to appear.

Which variables should be included in such an analysis? As mentioned in my research questions, it would be beneficial to exploit the current corpus for multiple sociolinguistic variables in order to determine whether an interlocutor effect is detected across variables, which would strengthen the argument for such an effect. In section 2.9.1, a review of previous research on L2 French included studies that examined multiple sociolinguistic variables from a single corpus (French & Beaulieu, 2016; Howard, 2012; Sax, 2003). French and Beaulieu examined *ne*-retention and /l/-deletion, hypothesizing that *ne*-retention is an easier sociolinguistic variable than /l/-deletion for learners to master; their results support this claim. Howard examined five variables (*ne*-retention, /l/-deletion, *on/nous*, liaison, and futurity) with the goal of establishing a sociolinguistic profile of the advanced learner; since acquisition of these variables does not take place in isolation, this sociolinguistic profile can determine which variables pose more, or less, difficulty for the individual learner. These variables encompass multiple linguistic domains (grammar and phonology) and differ as well in the markedness of their formal variant(s). Sax chose

to focus on four variables (/l/-deletion, *ne*-retention, interrogative structures, and *on/nous*) based upon their frequency of occurrence in informal speech (thus facilitating meaningful quantitative analysis) and the fact that they have been widely studied in previous research, allowing for comparisons across multiple learner proficiency groups.

Since my speakers are also distributed across learner groups at different proficiency levels, variables that appear in large numbers for speakers at each level (as well as for native speakers) would be ideal. This approach would disfavor some of the variables studied in previous research. For example, *on/nous* turns out to be nearly categorical in favor of the *on* variant in my native and near-native speakers. Only three Near-NSs used subject *nous* on more than one occasion: seven tokens (including one in quoted speech) from the near-native interlocutor (SaE) in Lille, four by speaker 1P, and nine by speaker 6P. One of the bilingual speakers (Th) used *nous* as a subject pronoun in one utterance; no other native speakers used subject *nous*. At the other end of the usage spectrum, as mentioned in section 3.6.3, four of the eight SA learners had categorical *nous* in variable contexts, significantly reducing the generalizability of a variationist analysis. The *on/nous* data certainly reveal some details about the sociolinguistic profile of speakers at different proficiency levels; however, a variable with nearly categorical variants in the majority of speakers is less useful in addressing questions of potential interlocutor effects.

Furthermore, variables that pose challenges concerning detection of their presence or absence may also be problematic. /l/-deletion can be relatively easy to detect in learners at lower proficiency levels, as Sax (2003) has noted, whereas with near-native and native speakers, the determination of presence or absence of /l/ is less straightforward. The automatic detection of liquids such as /l/ and /r/ using acoustic recognition software is much less advanced than detection

of nasality or of voicing, requiring judgments of [il] versus [i] to be made manually by the researcher.

I ultimately included the following variables for analysis: *ne*-retention and subject doubling. Based on previous studies, variable contexts for both of these structures are rather frequent in spontaneous oral production, with some corpora producing thousands of tokens of each variable context. *Ne*-retention and subject doubling can also appear in the same morphosyntactic "neighborhood"; previous studies (e.g., Villeneuve & Auger, 2013) have investigated how these two variables interact in L1 French, but no studies have, to my knowledge, investigated this interaction in L2 French. Furthermore, both variables also have a binary structure (retention or omission), with relatively straightforward detection of each variable even in rapid, spontaneous speech. This binary structure lends itself well to variationist analyses investigating the relative weights determining the probability of the appearance of each variant, which can allow us to identify the relative influence of potential interlocutor effects among all linguistic and extralinguistic factors, as discussed in section 2.10.

3.6.5 Statistical tools for variationist analysis

On a practical level, in order to analyze the effects of factor groups on a speech corpus containing, potentially, thousands of tokens, computerized statistical measures are necessary, and certain software programs have been created specifically for variable rule analyses in sociolinguistics. Goldvarb X (Sankoff, Tagliamonte & Smith, 2005) and Rbrul (Johnson, 2009) are two of the most common freely available computer software programs used in variationist studies. Both programs carry out a multivariate analysis which attempts to model the variation by finding the "best fit" through progressive iterations of the data. The "best fit" includes only the factors that are significant in accounting for the variation in the data, and these factors are also

ranked according to significance. For each group of factors, "factor weights" indicate how much a certain factor favors or disfavors the production of the variant under study when all factor groups are applied simultaneously. A factor weight of .5 indicates a neutral effect for the factor; a weighting greater than .5 indicates a favoring effect; and a weighting less than .5 indicates an inhibiting effect. Assuming proper identification of the factors that influence the variation, and assuming proper coding of each variant token, we obtain a probabilistic model for predicting which variant a particular speaker (or group of speakers) is likely to produce in a given context.

Currently, Rbrul allows for a more streamlined process of selecting and analyzing data compared with Goldvarb X. Rbrul can easily import data from spreadsheet programs such as Excel, which allows for the inclusion of meaningful factor group names and tokens (compared with Goldvarb's single-character identification of factor group names and tokens). Furthermore, Rbrul is able to more easily handle cases of factor groups containing a categorical variant for a specific factor ("knockouts"). Note that some recent studies (e.g., Donaldson, 2017) involving variationist analyses have employed logistic regressions rather than variable rule analyses such as those used by Goldvarb and Rbrul. Since logistical regression does not report output using factor weights, its results can be generalized more easily to disciplines outside sociolinguistics. However, output obtained from Goldvarb/Rbrul can be more easily compared to previous variationist studies. Given the ease of use of Rbrul, and the comparisons it allows with previous studies, I subsequently used Rbrul to analyze all of the data from the full-scale study, and the conclusions obtained from these variationist analyses were based on the results obtained from Rbrul.

3.7 Summary of methodology

This chapter has outlined the methodology undertaken to examine the research questions and hypotheses for the current study. After a discussion on the establishment of a formality

continuum in conversation through identification of informal features in French, I have outlined my research questions and hypotheses regarding the use of sociolinguistic variation and the effect of the interlocutor language background on learners of French. I have reported on an initial pilot study, which subsequently informed the methodology of the full-scale study. For the full-scale study, I have outlined the selection of participants and the procedures administered for each participant group, establishing a corpus of conversational French. I then analyzed the resulting corpus to determine that the conversations were indeed of an informal nature, allowing us to probe the question of how an interlocutor effect may influence the production of sociolinguistic variation. Finally, I have chosen two sociolinguistic variables for more detailed analyses, which will involve quantitative measurements and variable rule analyses.

The following two chapters will, in turn, examine these two variables: Chapter 4 will examine *ne*-retention and Chapter 5 will examine subject doubling. For each chapter, I begin with a motivation of the selection of these variables and a description of their morphosyntactic structures, followed by their treatment in previous studies in both L1 and L2 French. Then, for each variable I provide the results from the current study, including the results from variationist statistical analyses. I conclude each chapter with discussions on these results for each group of participants.

Chapter 4: *Ne*-retention

This chapter concerns the sociolinguistic variable of *ne*-retention. As initially discussed in Chapter 3, *ne*-retention is a potentially fruitful variable for making observations on factors influencing its use. Compared with other sociolinguistic variables in French, the high frequency of negation in conversational interactions, as well as the status of *ne*-retention as a binary variable (retention or omission), lends this structure particularly well to quantitative and variationist analyses. Though there can be some difficulty in determining the presence or absence of the *ne* particle in some cases in spoken French (Armstrong, 2002), such potentially ambiguous tokens are likely to affect only a small percentage of overall negation contexts.[21] Furthermore, *ne*-retention has a particularly significant sociolinguistic status in French—it is a "highly sensitive item in sociolinguistic terms" (Regan et al., 2009: 64), functioning as a stable variable expected in formal and written speech and, in informal speech, marking emphasis and contrast. Finally, we can directly compare these results with those found in Dewaele's 2004 study on *ne*-retention, the only other study on L2 French to treat the interlocutor L1 individual difference variable, and to the results in Donaldson (2017) for comparison with other Near-NSs. This section will briefly describe the basic structure of negation in French, followed by a review of studies on the *ne*-retention variable in L1 and L2 French.

4.1 Background on French negation

Verbal negation in Old French required a single marker of negation, *ne* (based on Latin *non*), which typically appeared pre-verbally, as in (1), with the Modern French equivalent in (2):

(1) *ne voil ublïer Bisclavret*
not want to forget Bisclavret

[21] For example, Donaldson (2017: 153) excluded 44 of 1,921 total negation tokens (2.3%) due to ambiguity in the input regarding the presence or absence of *ne*.

(2) *Je ne veux pas oublier Bisclavret*
I NEG want not to forget Bisclavret 'I do not want to forget Bisclavret'

This negation marker *ne* could be intensified by adding various modifiers, etymologically denoting small quantities, such as *pas* (lit. 'step'), *mie* (from *miette*, 'crumb'), *point* ('point'), and *goutte* ('drop'). The reinforcement of emphatic *pas* with the semantically negative marker *ne* was already present in surviving texts of the 12th-century epic poem *La Chanson de Roland* (cf. Pohl, 1975), and over time, the emphatic forms began to acquire the meaning of negation, in part due to intonation patterns placing phrase-final stress on these post-verbal items (cf. Dewaele, 2004a; Ludicke, 1982). Thus, the *ne* particle was weakened to the point of functional obsolescence as several emphatic forms became grammaticalized as verbal negation markers, a process that was essentially complete by the end of the 17th century (Martineau & Mougeon, 2003). This type of transformational sequence of negative expressions has been termed the "Jespersen Cycle," a process outlined by the Danish linguist Otto Jespersen (though initial observations on this sequence by other linguists such as Meillet and Blancquaert preceded Jespersen's). Jespersen (1917: 4) described this phenomenon in various languages as follows: "[T]he original negative adverb is first weakened, then found insufficient and therefore strengthened, generally through some additional word, and this in its turn may be felt as the negative proper and may then in course of time be subject to the same development as the original word." Analogues can also be found in the evolution of English (*ic ne secge* → *I ne seye not* → *I say not*, 'I don't say') and German (*nisagu* → *ih ensage niht* → *ich sage nicht*, 'I don't say').[22] In the Gallo-Romance dialect that became the basis for Modern French, post-verbal *pas* became the primary marker of negation, and pre-verbal

[22] Labelle (2019) contends that the term "cycle" is a misnomer, at least for French, given that the marker of clausal negation at the end of the cycle (viz., *pas* in contemporary French) does not have the same properties as the initial marker (*ne* in medieval French). She proposes an updated characterization of the evolution of negation as "spiral-like rather than cyclic, since the endpoint is distinct from the initial point" (p. 159).

ne was retained as bipartite negation became standard in the rapidly expanding written forms of the language,[23] while *ne* began to be eliminated in spoken forms as early as the 16th century.[24] Elsewhere, Coveney (2002: 62) notes that *ne* seems to have dropped out of Occitan and Canadian French rather quickly due to a relative paucity of written norms. However, standard Modern French has retained bipartite negation in formal and written styles and, due to its high level of prestige, contemporary native and non-native speakers tend to perceive bipartite negation as the "correct" form, even though *ne*-deletion is almost categorical in certain spoken styles. Coveney (2002) also notes that stigmatization of *ne*-deletion in spoken L1 French does still exist but may largely be limited to attitudes of teachers toward students (cf. Lafontaine, 1986: 126). Nevertheless, as noted by Fagyal, Kibbee, and Jenkins (2006), *ne* remains an important factor for demonstrating communicative competence in certain Francophone communities; for the French language entrance exam at the Université du Québec en Outaouais, for example, *ne*-retention is still considered one of the criteria for judging whether a student has sufficient command of French syntax in written form, despite empirical observations of near-categorical *ne*-deletion in spoken Canadian French.

[23] Modern French still contains examples where *ne* alone can carry the semantic content of negation of verbs such as *savoir, cesser*, *pouvoir*, and *oser*, typically when such negated verbs are followed by a non-finite verb. In these cases, negation of these verbs with *ne* expresses identical semantic meaning whether or not *pas* is present.

Nous ne pouvons partir.
Nous ne pouvons pas partir. 'We cannot go.'

[24] Martineau and Mougeon (2003) cite evidence for *ne*-deletion in the early 17th century based on the Héroard diaries, which documented the child speech of Louis XIII from 1605 to 1611, indicating that *ne*-deletion was already prevalent, at least in children's speech, by this time. However, Labelle (2019) contends that such children's speech represents acquisition of a variable rule of *ne*-deletion rather than incomplete acquisition of bipartite negation; citing Ayres-Bennett's (1994) examples of *ne*-deletion in transcription of *adult* speech in the Héroard diaries, Labelle posits that *ne*-deletion must have begun as early as the 16th century).

4.2 Studies on *ne*-retention in L1 French

Contemporary studies on *ne*-retention in oral production have been carried out since the mid-20th century. Since the first empirical observations were made in the 1950s, *ne*-retention appears to be broadly declining in overall use, but there is still much inter-speaker variation, and *ne*-retention percentages can be considerably influenced by the social profile of speakers chosen, the method of data collection (e.g., interviews versus undirected conversation, or telephone versus face-to-face communication), the location of data collection, and the choices made concerning the inclusion or exclusion of invariable or marginally invariable structures (e.g., whether to exclude lexicalized formulae). Concerning the location of data collection, early studies provided evidence for geographical variation concerning *ne*-retention patterns in France (Lüdicke, 1982; Pohl, 1968), and, as introduced at the end of the previous section, it is well known that Québec French has demonstrated near categorical *ne*-deletion since at least the 1970s (Sankoff & Vincent, 1977).[25] There is evidence that Swiss French is also approaching categorical deletion (Fonseca-Greber, 2007; Meisner, 2016). Importantly, and especially with regard to L2 French, even though *ne*-retention may still be regarded as a more prestigious variant in some Francophone communities (or subsets of speakers in these communities), omission of *ne* in informal styles is not subject to stigmatization to the same extent as other sociolinguistically conditioned variables such as interrogative structures.

Table 4-1 provides an overview of corpus studies conducted on *ne*-retention in L1 French, listed by year of data collection for European French followed by Canadian French, and including total verbal negation tokens with overall *ne*-retention percentages. When available, ranges of inter-speaker variation are provided, following Donaldson's (2017) observation that overall averages of

[25] Though, as Coveney (2002: 90, citing Lemieux 1985: 101) points out, *ne* should still be considered as part of the grammars of these speakers, as they are capable of producing it in more formal styles.

speakers in a study often mask considerable inter-speaker variation. Other information on speakers is also provided where available. Many authors report data in terms of *ne*-deletion, though in more recent work (e.g., Armstrong, 2002; Ashby, 2001; Coveney, 2002; Donaldson, 2017; Fonseca-Greber, 2007; Hanson & Malderez, 2004) the data have been expressed in terms of *ne*-retention. As mentioned in Chapter 2, I will report figures in terms of *ne*-retention for the studies cited here (converting figures from deletion to retention as necessary) as well as for data obtained in the current study.

Table 4-1. Studies on ne-retention in L1 French (adapted from Auger & Villeneuve, 2008; Donaldson, 2017)

Study	Year of survey	Research site(s)	Speakers	Total negation tokens	*ne*-retention: overall %	*ne*-retention: range	Other information
Zwanenburg (1965)	1959	France	10	99	29.3%	N/A	Ages: 25-65
Pohl (1975)	1950s-60s	Saint-Mard (Belgium); Paris; elsewhere in France	More than 25	5,308	61.9%	13.3 – 97.8%	Collection of oral data from various sources[26]
Ashby (1976)	1967-68	Paris	50	1,029	55.8%	8 – 100%	Middle-class Parisian adults
Diller (1983)	1975	Béarn (Pyrénées-Atlantiques)	12	641	65.7%	N/A	Older adults in southern rural France
Ashby (1981)	1976	Tours	35	2,818	36.6%	0 – 93%	Ages: 14-21 & 51-64
Coveney (2002)	1980	Somme	30	2,932	18.8%	0 – 54.8%[27]	
Lüdicke (1982)	N/A	France (various regions)	5	N/A	N/A	8.7 – 87.1%	5 females from all ages/social classes
Moreau (1986)	1982-83	Belgium	30	3,158	50.3%	3 – 96%	Radio transcripts of male French celebrities
Pooley (1996)	1983	Roubaix (Nord)	61	3,719	7.0%	N/A	
Hansen & Malderez (2004)	1989-93	Paris; Oise	48	1,329	8.2%	0 – 56%	
Armstrong (2002)	1990	Dieuze (Lorraine)	16	2,501	1.9%	0 – 8.3%	Ages: 11-19
Pooley (1996)	1995	Rouge-Barres (Nord)	15	391	1.0%	N/A	Adolescents
Ashby (2001)	1995	Tours	30	1,593	15.7%	0 – 70%	
Fonseca-Greber (2007)	late 1990s	French-speaking Switzerland	14	1,982	2.5%	N/A	
Culbertson (2010)	2002-05	Lyon	5	5,990	7.6%	0.7 – 12.5%	Child-directed speech from adult caregivers[28]
Auger & Villeneuve (2008)	c. 2005	Vimeu (Somme)	6	711	20.8%	7.1 – 44.2%	Mostly older males; 3 bilingual French-Picard
Donaldson (2017)	2006	Paris; Pyrénées-Atlantiques	10	985	11.0%	0 – 36.5%	Interlocutors = near-native speakers of French

[26] The sources cited by Pohl include the *Français fondamental* corpus (Gougenheim, Rivenc, Michéa, & Sauvageot, 1964) and the study by Zwanenburg (1965), which had prosody as its focus but included full transcriptions of spoken language data, from which *ne*-retention could be analyzed.

[27] For various reasons, Coveney advances an argument for considering the speaker with 54.8% *ne*-retention as an outlier; the speaker with the second-highest *ne*-retention is at 46.9%. *Ne*-retention rates for all of Coveney's speakers can be found in the appendix to his article on subject doubling (Coveney, 2005: 109).

[28] From the Lyon Corpus (Demuth & Tremblay, 2008).

Table 4-1 (continued). Studies on ne-retention in L1 French (adapted from Auger & Villeneuve, 2008; Donaldson, 2017)

Study	Year of survey	Research site(s)	Speakers	Total negation tokens	*ne*-retention: overall %	*ne*-retention: range	Other information
Villeneuve & Auger (2013)	2006-07	Vimeu (Somme)	8	457	31.9%[29]	12 – 67%	Ages: 30-75; all males; 4 bilingual French-Picard
Torreira, Abba-Decker & Ernestus (2010)	2007	Paris (Île-de-France)	46	±10,000	6.7%	N/A	Nijmegen Corpus of Casual French; 23 pairs of speakers + confederate
Palasis (2015)	2007-08	Southeastern France	19	1,098	1.8%	N/A	Kindergarten children; age range 3.5 – 5
Spithoff (2018)	2008	Orléans	N/A	11,649	10.5%	N/A	ESLO2 (Orleans) Corpus
Meisner (2016)	2008-09	Paris (Île-de-France); Neuchâtel, Switzerland	54	334	17.7%	0 – 100%	T-zéro corpus; 2-5 minute recordings
Stark (2012)	2009-11	Switzerland	N/A	1330	23.5%	N/A	Text messages; SMS4science.ch corpus
Canadian French							
Sankoff & Vincent (1977)	1971	Montréal	60	±10,000	0.5%	0 – 8%	from Sankoff-Cedergren (1972) corpus
Poplack & St-Amand (2007)	early 1980s	Ottawa-Hull	120	61,316	0.2%	N/A	Ottawa-Hull corpus

[29] The authors considered certain collocations (e.g., *il y a, c'est, il faut*) as near-categorical regarding *ne*-deletion and did not include them in the results.

As Table 4-1 shows, an analysis of the range of inter-speaker variation can provide some context to the overall *ne*-retention rates obtained in these studies. Even in early studies from the 1970s, *ne*-retention for some speakers was below 10%, and no studies besides Ashby (1976) observed any speaker at 100% retention. Moreover, a majority of the studies reported here include at least some speakers with categorical *ne*-deletion, beginning in the 1970s and continuing to the present, representing several regions of France in addition to Canada (where audio recordings show near-categorical deletion beginning at least by the 19th century; cf. Poplack & St-Amand, 2007).

4.2.1 Factors influencing *ne*-retention in L1 French

It is clear that an analysis of *ne*-retention must examine demographic factors in the population of speakers studied, especially in light of the numerous studies with widely ranging inter-speaker variation. Why would one speaker have nearly categorical *ne*-retention while another speaker in the same study (and same research site) has categorical deletion? A review of the literature provides evidence that both linguistic and extralinguistic factors interact with this variable with varying degrees of influence.

Ashby (1976) was among the first to merge both linguistic and extralinguistic factors into a single analysis, identifying those that favored *ne*-retention (see Table 4-2). Ashby notes that the first seven linguistic factors had a more significant effect on *ne*-retention than the remaining five.

Table 4-2. Factors that favor ne-retention in Ashby (1976)

Linguistic	**Extralinguistic**
Second negative[30] other than *pas*	Female speaker
Reinforcing adverb (e.g., *Je ne veux absolument rien*)	Administrators/other professionals
Full NP rather than subject clitics	Older speakers
Subject clitics *nous/vous*	First half of conversation
Dependent clauses	Narrative or explanation
Verbal mood other than indicative	
Less frequent expressions (e.g., lack of formulae such as /ʃepa/ for *Je (ne) sais pas*)	
Third negative (e.g., *Je ne fais jamais rien*)	
Lack of adverb *non* (e.g., *Non, je ne l'ai pas vu*)	
Slow speech rate	
Intervocalic position	
Formal pronoun of address (*vous*)	

Ashby conducted a *chi*-square test for each of these factors individually. Though his study did not account for multiple and simultaneous factors that variationist analyses can identify (as outlined in section 2.10), it established that even among more conservative French speakers (Parisian upper-middle class), *ne*-deletion is common in informal speech.

In other studies, social class (cf. Lüdicke, 1982) and age (cf. Coveney, 2002) have appeared as primary extralinguistic factors, with some influence from gender (cf. Ashby, 2001; as is borne out in other sociolinguistic studies, female speakers gravitate toward more formal styles, resulting in this case in higher *ne*-retention). Part of the explanation for age as a significant factor may be due to age-grading influence rather than to wholesale shifts in each age group over time (Coveney, 2002: 90): as youths, native speaker speech almost completely lacks *ne*, despite evidence of higher *ne*-retention in child-directed speech[31]); as adults, with pressure from professional environments and from the written language as standard, their use of *ne* increases; and finally, as older speakers,

[30] Ashby (1976) specifies *ne* as the "first negative" and *pas/jamais/rien/etc.* as the "second negative."

[31] See Culbertson (2010) for more discussion on *ne*-retention in child-directed speech; see Clark and de Marneffe (2012) and Dye (2011) for more discussion on the verbal domain in child-directed speech.

ne-retention decreases due to less professional pressure and due to a return to more informal interaction with the community. This "U-shaped curve" is by no means identical in all speakers, as social networks, individual motivation, and other sociolinguistic factors all intervene to varying degrees. In a cross-sectional sample, Meisner (2016) found that *ne*-retention was actually higher in youths aged 14-18 (21%) than in young adults aged 19-24 (5%) and 25-34 (16%); between the ages of 30 and 40, however, *ne*-retention picks up again and increases from 50% in the age 35-44 group to 67% in both the age 45-54 and 55-64 group.

At least one study (Stark, 2012) suggests that contemporary use of *ne* is no longer influenced by sociolinguistic factors, given the currently low retention rates and the apparent leveling across some of these factors, such as age and regional differences. Donaldson (2017: 160) counters that *ne* at least carries "sociostylistic value" and provides evidence of increased *ne*-retention in specific extralinguistic situations: to mark emphasis, to quote speech from a more formal register, and to signal serious topics. Fonseca-Greber (2007) determined emphatic negation to be one of the few remaining contexts for *ne*-retention in Colloquial Swiss French; emphatic negation may include lexical emphasis, repeated speech, slower speech, pitch prominence, and contrast, all of which may favor the retention of *ne*. Fonseca-Greber (2007: 267) also stresses that the determination of an utterance as emphatic is not necessarily a binary distinction: "It is perhaps better to think of emphasis not as [+Emph.] or [-Emph.] but as a continuum with degrees of emphasis…where the more emphatics co-occur, the more emphatic the utterance." Her examples and explanations of each type of emphatic negation in her corpus also involve a certain degree of subjective determination concerning whether the utterance meets a threshold for consideration as "emphatic." Often, an utterance contains multiple lexical elements clearly indicating emphasis,

such as in (3), where a post-verbal intensifying adverb (*strictement*) and type of negator (*aucune*) emphasize the negation.

(3) *l'apprenti n'avait strictement aucune idée*
'the apprentice had absolutely no idea'

For other emphatic uses, such as slower speech, Ashby (1976) uses the syllabic rate of the utterance as measured in syllables per minute for determining contexts of slower speech. Fonseca-Greber (2007: 263) adopts a more subjective approach by simply noting negation contexts in which the speaker's speech rate was noticeably slower than surrounding utterances—"breaking the normal rhythm" in order to draw her interlocutor's attention to the content of the particular utterance.

Focusing now on purely linguistic factors conditioning the presence or absence of *ne*, many studies have built on Ashby's (1976) findings. Certain factors appear repeatedly, while the influence of other factors has been less clearly established in cross-sample comparisons. In nearly every study on *ne*-retention, the type of post-verbal negator has been found to influence the use of *ne*, generally finding that negators other than *pas*, such as *rien, personne, plus,* restrictive *que*, and *aucun*, favor *ne*-retention. The addition of a second negator (e.g., *je ne vois plus personne* 'I no longer see anyone anymore') has also been shown to favor *ne*-retention (Pooley, 1996). However, since these post-verbal negators appear much more infrequently than *pas*, their influence has been less robustly identified, in certain cases producing generalizations based on very small numbers of tokens. Overall, it appears that *pas* favors the lowest *ne*-retention (Ashby, 1981; Coveney, 2002; Donaldson, 2017; Meisner, 2016; Pooley, 1996).

Concerning the nature of the subject, certain personal pronouns favor *ne*-deletion, such as *je* and *il* (Ashby, 1976; Coveney, 1996; Armstrong & Smith, 2002) and *ce/c'/ça* and *on* (Pooley, 1996: 173), with impersonal *il* (e.g., *il (ne) faut pas* 'one must not') favoring deletion over personal *il* (e.g., *il (ne) vient pas* 'he's not coming') (Pooley, 1996). Personal pronouns in general favor *ne*-

deletion over nominal expressions (Ashby, 1981; Coveney, 2002; Diller, 1983; Hansen & Malderez, 2004). Meisner and Pomino (2014) also find that full DPs and "heavy" pronouns (*nous, vous*, and *elle*) favor *ne*-retention. The lack of subject, as in negated infinitives, favors *ne*-retention (e.g., *C'est difficile de ne pas être d'accord*, 'It's difficult to not agree'), as does the relative pronoun *qui*, functioning as a subject of a relative clause (Coveney, 2002; Hansen & Malderez, 2004; see also Hirschbühler & Labelle, 1994, for an extensive overview of negated infinitives). As far as a scale of favorability to *ne*-retention concerning subjects, Meisner (2016) situates relative *qui* between proper nouns and lexical nouns on one side (favoring *ne*), and clitics and demonstrative *ça* on the other (disfavoring *ne*). Doubled subjects are also likely to favor deletion (Auger & Villeneuve, 2008); I will examine this construction in detail in Chapter 6.

Phonological environment has also been identified as an influence on *ne*-retention, with Ashby (1976) finding intervocalic position to favor retention. Meisner (2016) also found this environment to favor retention, though in a small sample size (60% retention in 15 tokens compared to 18% retention in 334 tokens overall). From the same corpus, Meisner and Pomino (2014) find evidence that *ne* is sensitive to the phonological form of the preceding element; unstressable, bi-segmental subjects (*je, tu, il/ils, on, ce*) favor deletion while stressable subjects favor retention.

The nature of the verb can play a role in affecting *ne*-retention, as shown by Moreau (1986) and Gadet (1997); frequent verbs such as *être* and *avoir*, and modal verbs such as *devoir* and *pouvoir*, favor *ne*-deletion, though Meisner (2016) found this effect to be marginal. Some frequent verbs are used in common expressions such as *c'est pas* or *il y a pas*, and these have been considered as lexicalized expressions or "preformed sequences" (Moreau, 1986) that favor *ne*-deletion (Ashby, 1981; Coveney, 2002), to the extent that in some studies (e.g., Villeneuve &

Auger, 2013), such sequences are excluded from variationist analyses of *ne*. Ashby (1976, 1981) also provided early evidence that the type of clause containing the negated verb can affect *ne*-retention; subordinate (versus main) clauses and transitive (versus intransitive) verbs favored *ne*-retention, a result supported by Sturm (1981) and more recently by Meisner (2016).

Intervening elements can also influence the presence or absence of *ne*. Between the subject and the verb, the presence of object clitics can contribute to omission of *ne*, such as in *je ne le lui ai pas donné* ('I didn't give it to him/her'), where *ne* is the weakest element in the clitic sequence, semantically superfluous, and is likely to be dropped in spontaneous, rapid speech for reasons of phonotactic simplicity (cf., e.g., Posner, 1985; Larrivée, 2014; Meisner, 2016). Between the verb and the negator, there is some evidence of a reinforcing adverb favoring *ne*-deletion (Ashby, 1976; Meisner, 2016), though other studies (e.g., Donaldson, 2017) did not find a significant effect, in part due to small sample sizes. Hansen and Malderez (2004) and Meisner (2016) did find a correlation between *ne*-deletion and presence of a reinforcing adverb when following the negator (e.g., *je n'aime pas <u>vraiment</u>* 'I don't really like').

Finally, Coveney (1998) and Donaldson (2017), among others, have pointed out the difficulty in providing clear evidence for the individual influence of each of these linguistic factors. Different methodologies across studies can make cross-sample comparisons less robust, and, as mentioned above, small sample sizes for some factors render their influence less generalizable.

In his early study on *ne*-retention in native French speakers, Pohl (1975: 25) concludes with the subjective statement that "there are cases where speaking French too well [i.e., categorical *ne*-retention] may mean speaking it badly." In other words, perceptions of what constitutes speaking "correct" French do not always align with expectations of language use for a given sociolinguistic context. Pohl also posits that the expectation of deleting *ne* in certain styles in

contemporary French has become so "normal" that anyone who always retains *ne* might be taken for a "foreigner." Though his analysis did not treat L2 French speakers, this stereotype of L2 speech supports the arguments outlined in Chapter 2, where L2 speakers demonstrate more "classroom-like" speech patterns in interactions with the target language community, and in the case of *ne*-retention, the stereotype presupposes that L2 speakers never (or nearly never) drop *ne*. This position forms a good starting point for discussion of *ne*-deletion in L2 French, as developed in the next section.

4.3 *Ne*-retention in L2 French

I begin with a brief sketch of typical classroom instruction of verbal negation in L2 French, followed by the typical developmental sequence of learners. Most classroom instruction of French presents negation as a bipartite structure (*ne…pas*), and textbook input overwhelmingly favors *ne*-retention. Etienne and Sax's (2009) survey found that nine of 14 introductory textbooks, and only three of eight intermediate textbooks, instruct learners that *ne* may be omitted in informal speech. Generally, in (constructed) informal dialogues that include informal interrogative variants and slang or colloquial speech, *ne* is consistently maintained in verbal negation, which in some cases creates an "involuntary comic effect," such as the oral activity in one textbook where learners are to role-play an argument between husband and wife with the title "*Je n'en peux plus*" ('I can't take it anymore')—what Etienne and Sax (2009: 593) describe as a "potentially great opportunity to practice stylistic variation" that is simply ignored in favor of standard (if unrealistic) *ne*-retention.[32] Mougeon et al. (2002) also report on a textbook for L2 French learners in which the only characters

[32] It may, however, be noteworthy to mention that retaining *ne* in the constructed argument dialogue could be considered an example of "microstyle" variation, where *ne* is retained to signify emphasis on the negated aspect of the sentence, or to convey strong emotions. In the classroom, this type of emphasis would likely not be mentioned as an exception to the trend toward *ne*-deletion in informal settings.

to omit *ne* were negatively portrayed. As for input from classroom instructors concerning *ne*, recall the observation by Rehner and Mougeon (1999) that even in immersion settings, due to the typically higher formality of the classroom environment, teachers are likely to produce high levels of *ne*-retention, though in "unguarded speech" teachers may omit *ne*, providing indirect clues to learners as to the socio-stylistic evaluation of *ne*-retention.

Teachers of French often note that in the first stage of acquisition of verbal negation, learners sometimes omit the negator (*pas*) while retaining the semantically empty *ne*, as in *Je ne vais au parc* ('I NEG go to the park'). A variety of factors may cause learners to produce this structure. Processing may cause difficulties in the production of two morphemes of negation, so that learners only produce one morpheme and deem the negation complete after producing *ne* (especially in syntactically more complex VPs such as in *passé composé* and modal verb constructions). In addition, transfer from the L1 may play a role, such as in English and Spanish, languages in which only one morpheme is required for verbal negation and in which verbal negation bears orthographic similarities (*no, not*) to *ne.* Following this stage, learners typically produce both elements of the standard bipartite negation, though due to its high frequency, *pas* is often overgeneralized and produced in place of, or in addition to, other negative quantifiers (cf. *Personne ne veut (pas) manger*; 'No one NEG wants (not) to eat'). The final, optional stage in this "long circuitous route" (Dewaele, 2007: 6) of acquisition of French negation involves the sociolinguistically conditioned omission of *ne* through either explicit instruction or input from other speakers; as discussed in Chapter 2, learners usually retain *ne* at non-nativelike levels unless they have had extensive exposure to native speakers and/or a lengthy residency in a French-speaking community.

Due to its sociolinguistically marked status in informal contexts, frequent opportunities for verbal negation in written and spoken language, and large differences in retention rates between NSs and NNSs, *ne*-retention in L2 French has been the subject of a number of sociolinguistic studies, the first dating to the early 1980s. Across a range of methodologies including various data elicitation techniques, these studies have shown considerable variation in *ne*-retention rates across groups of learners, ranging from 11% in Thibault and Sankoff's (1997) study to 100% in a beginner group in Sanell's (2007) study, as well as extreme inter-individual ranges for studies reporting these ranges, such as those observed in Trévise and Noyau's (1984) learners (range of 1-100% *ne*-retention), or Sax's (2003) advanced learner group with long-term study-abroad (range of 2-92% retention). Table 4-3 reports these studies on pp. 129-130, organized by date of publication.

Table 4-3. Studies on ne-retention in L2 French

Study	type of study	Research site(s)[33]	# of learners	Learner L1	L2 Proficiency; years of study	L2 environment	Discourse elicitation format	Interlocutor status	Total tokens	*ne*-retention: overall %	*ne*-retention: range
Painchaud et al. (1982)	longitudinal	Montréal	36	various Southeast Asian	varied[34]	living abroad	interview	NS	1788	Overall: 59% T1: 67% T2: 51%	N/A
Trévise & Noyau (1984)	cross-sectional	France	8	Spanish	LOR range: 0.5 to 13 years; average = 5.3	living abroad	interview	NS	N/A[35]	42%	1 – 100%
Regan (1996)	longitudinal	France; Brussels (Belgium)	6	(Irish) English	advanced; 6 years	SA (study abroad) (1 year)	interview	NNS=1	762	Overall: 49.3% Pre-SA: 52% Post-SA: 48%	81 – 92%
Thibault & Sankoff (1997)[36]	cross-sectional	Montréal	17	English	varied	(classroom) immersion	interview	NS=2; NNS=1	2283	11%	N/A
Rehner & Mougeon (1999)	cross-sectional	Toronto (Ontario)	40	English	Grade 9-12; 0-4 years	classroom immersion	interview	NS=1 (European French)	2163	72%	N/A
Dewaele & Regan (2002)	cross-sectional	Brussels (Belgium)	27	(Flemish) Dutch	advanced; 4-6 years	classroom instruction	formal/ informal interviews	NS=1 (bilingual Fr-Dutch)	992	81%	2 learners at 100%
Sax (2003)	cross-sectional	varied:	30 total[37]	English	varied:	varied:	formal/ informal role plays	NS=2	3749	60.8% overall	0.7 – 100%
		US	14		intermediate/advanced; 2-4 years	classroom instruction			1172	96.8%	67.6 – 100%
		US/France	6		advanced; 4+ years	classroom instruction/ SA (2-6 mo)			831	75.6%	35.3 – 92.6%
		US/France	10		advanced; 4+ years	SA (8.5 mo – 4 yrs)			1746	29.7%	0.7 – 91.9%

[33] The research site applies to the location where learners were exposed to primary input in the target language, even if data elicitation took place in the L1 community.

[34] Fifteen of 36 participants reported some study of French at school; all were studied before and after finishing an intensive 30-week French learning program (900 total instruction hours) for new immigrants.

[35] The authors reported total negative sentences for only one subject (n = 202); with eight total subjects, the number of total negation tokens can be presumed to be at least 1000.

[36] From the corpus obtained by Sankoff et al. (2017).

Table 4-3 (continued). Studies on ne-retention in L2 French

Study	type of study	Research site(s)	# of learners	Learner L1	L2 Proficiency; years of study	L2 environ-ment	Discourse elicitation format	Interloc-utor status	Total tokens	*ne*-retention: overall %	*ne*-retention: range
Dewaele (2004a)	cross-sectional	U.K.	64	various; mostly English	pre-advanced to advanced; 5-11 years	classroom instruction/ SA	formal/ informal interview	NNS = 55; NS=9	N/A	72.8%	11 – 100%
Thomas (2004)	longi-tudinal	varied:	87 total	English	3-4 years	varied:	re-tells and read-aloud	none	1365	72.1% overall	N/A
		France	48			SA (1 year)				Pre-SA: 78.7% Post-SA: 72.7%	
		Ontario	39			classroom instruction				Pre: 68% Post: 80.3%	
Sanell (2007)	cross-sectional	various (Interfra corpus)	24	Swedish	beginner to near-native	various	interviews	NS	1275	Overall: 41.8% Beginner: 100% Intermediate: 62% Adv low, mid, high: 51%, 33%, 31% Near-native: 11%	0 – 100%
Howard (2012)	longi-tudinal	France	5	English	7 yrs	classroom instruction	informal interview	NNS	at least 181[38]	at least 42%	10 – 100%
					8 yrs (1 yr study abroad)	study abroad			273	26.4%	2.3 – 37.5%
					9 yrs (1 yr study abroad)	classroom instruction			340	31.5%	6.9 – 60.9%
van Compernolle & Williams (2012b)	longi-tudinal	US	24	English	1 yr	classroom instruction	online chat discussions (CMC)	NNS (some NS)	357	Overall: 87.9% Week 2: 98.5% Week 13: 90%	72.4 – 100%
French & Beaulieu (2016)	longi-tudinal	Québec	18	various	advanced	immersion/ SA	planned/ unplanned re-tells	none	114	Planned T1: 86%, T2: 10%; Unplanned T1: 10%, T2: 58%	N/A
Donaldson (2017)	cross-sectional	France	10	English	near-native	living abroad	informal conversation	NS	892	22.4%	4.8 – 38.0%

[37] Sax divides her subjects by factor group, so the groupings by study-abroad experience do not necessarily correspond to years of classroom study. Since other studies (e.g., Thomas 2004) group learners by study-abroad experience, this is how the learner groups in Sax's data will be presented.

[38] Total negation tokens in this column were not published but were determined via the author's reported percentages; tokens for learners with categorical *ne*-retention were not published.

As Table 4-3 demonstrates, there is great variability not only in the study design (e.g., data elicitation format, cross-sectional versus longitudinal comparisons), but also in the selection of speakers (learner L1, learner proficiency) and, as discussed in Chapter 2, the choice of interlocutor. Unlike in L1 French studies, there does not seem to be a diachronic trend toward less *ne*-retention in more recent studies, since learners are being analyzed at various stages of their acquisition. What is clear is that proficiency and time spent abroad play a large role in *ne*-retention rates, as discussed in Chapter 2 and elaborated in the next subsection. While it is difficult to make clear comparisons across learner groups and take into account the intensity and length of target language exposure, learners at the lowest proficiency levels in these studies tend to retain *ne* at rates over 90% (Sax, 2003; van Compernolle & Williams, 2012b). These speakers appear to be close to the minimum threshold for holding a conversation or producing narration in L2 French.

As one may expect, study abroad can impact learner use of *ne*-retention. Longitudinal studies such as those by Howard (2012) show that speakers with extensive classroom experience (at least 7 years) still retain *ne* at high rates (with some speakers still producing categorical retention); after a year-long study abroad, all five learners had lower *ne*-retention. Sax (2003) also found a clear difference between two groups of advanced learners, one of which had had a long-term study abroad (30%) and the other a short-term study abroad (76%).[39] Despite this evidence of convergence with nativelike *ne*-retention patterns, sojourns abroad are not necessarily a guarantee for more targetlike *ne*-retention, as Thomas (2004) demonstrated with pre-study-abroad learners at 79% retention but post-study-abroad (one year) maintaining at 73% retention.

[39] Grouping speakers by other categories of proficiency, Sax also found significant differences in students enrolled in a graduate program in French (39% *ne*-retention overall) compared with groups of undergraduate students enrolled in second-year and fourth-year French courses (99% and 59% *ne*-retention, respectively)

Furthermore, Howard (2012) found that, with a year of classroom instruction after study abroad, *ne*-retention increased slightly (from 26% to 32% overall).

Perhaps unsurprisingly, near-native speakers living in target language communities produce *ne* at the lowest overall rates, such as those found in Donaldson (2017) with minimum five years residency in France, or Anglophone Montrealers in Thibault and Sankoff (1997) in daily contact with Québec French speakers whose native *ne*-retention approaches 0%. However, as in studies on L1 French, overall rates can mask considerable ranges across individual speakers. Trévise and Noyau (1984) found that learners with long residence in a French-speaking community (Paris) deleted *ne* at rates exceeding many native speakers; two speakers with 13 years residency had 1% and 10% *ne*-retention in informal interviews. Donaldson (2017) also had three speakers at less than 10% retention, while Sax (2003) had one advanced learner at less than 1% retention and several speakers in the 14-20% range.

4.3.1 Factors influencing *ne*-retention in L2 French

Concerning specific linguistic and extralinguistic factors affecting *ne*-retention in L2 French, results vary by study, but most reveal some version of target language exposure as significant. Rehner and Mougeon (1999) found two factors—language spoken at home and non-*pas* negators (e.g., *rien, jamais*)—as significant in producing more deletion; contact with a Francophone environment was also important. Thomas (2004) also found that intensive contact with Francophones in the target community was a key factor, while Dewaele (2004a) found that personality (e.g. extraversion) and frequency of French use were significant factors. Donaldson (2017) also demonstrated that more advanced L2 speakers may show sensitivity to sociostylistic variation as native speakers do, where, beyond constraints due to purely linguistic factors, *ne* may

be retained in an informal situation due to serious topics, emphatic negation, or quoting from a more formal register.

Results from these studies generally support the conclusion that, for students to start omitting *ne*, they need either explicit instruction or opportunities for authentic interaction with NSs, and for speakers to produce more nativelike *ne*-retention rates as well as more nativelike sensitivity to linguistic and extralinguistic factors, interaction with the target language community is crucial. Even with target language input, however, *ne*-retention in advanced learners may be subject to "fossilization" at a non-nativelike level. Based on previous studies (Dewaele & Regan, 2002; Regan et al., 2009; Sax, 2003), Donaldson (2011a) initially speculated that *ne*-retention rates in L2 French speakers would never reach nativelike levels, though he subsequently observed that even for highly proficient learners, "intensive interaction and active commitments appear a necessary but not solely sufficient condition for acquiring nativelike rates of *ne*" (Donaldson, 2017: 164).

It must again be emphasized that, while lower *ne*-retention generally corresponds more closely to native speaker patterns, learners wishing to integrate into the target language community should not "aim" for 0% *ne*-retention, as this would be an over-correction for the general tendency to delete *ne* in informal situations (cf. Regan, 1996). Rather, learners wishing to speak as much like native speakers as possible would need to navigate the specific situations in which *ne* would be retained in informal situations. Given all the possible linguistic and extralinguistic constraints on such a sociolinguistic variable, for most learners this "minefield" navigation generally involves a period of "unlearning" classroom (and prescriptivist) norms—or at least acquiring more experience with non-written forms of the language—and identifying the situations in which *ne* is typically omitted by members of the target language community. This process may never actually

finish; learners may either continually adapt their speech based on input they receive, or they may adopt a certain style that may not pattern like native speakers (with either comparatively too little or too much *ne*-retention) but is considered by the learner to be sufficient for successful communication with her community of speakers (recall, from discussion in Chapter 2, van Compernolle's (2015) characterization of learners as cautious outsiders with respect to sociolinguistic variation and Donaldson's (2017) discussion on non-native pragmatic conservatism).

Given the sometimes large ranges obtained in native speaker *ne*-retention studies, comparing raw frequencies between native and non-native speakers cannot paint the whole picture for this sociolinguistic variable. Crucially, the process of acquiring sociolinguistic competence also involves modification of the L2 grammar with respect to the factors that may favor or disfavor production of the sociolinguistic variable, in this case *ne*-retention. Therefore, in reporting *ne*-retention results from the current study, I begin data analysis from an overall perspective of *ne*-retention rates, followed by specific factors that appear to influence *ne*-retention based on a variationist account of the data.

4.4 Results: *ne*-retention

From my corpus of 59 conversations involving SA learners, Near-NSs, and all interlocutors, all utterances containing verbal negation were analyzed for retention/omission of the *ne* particle. I excluded cases of phonological ambiguity (e.g., *On (n')est pas content* 'One (NEG) is not happy') as well as verbal negation contexts in which the subject was not indicated (e.g., *faut pas exagérer* '[One] must not exaggerate'). Primarily concerning SA learners, I also excluded instances of *ne* lacking the post-verbal negator, such as *il n'habite avec lui* ('he NEG lives with him'), per Sax (2003: 107). In her study, Sax found several dozen examples of *ne* without a negator

(54 of 3749 total tokens, 1.4%); in the current study, a similar percentage of SA learners produced *ne* with no negator (6 of 500 total tokens, 1.2%). Instances of false starts or self-correction in all speakers where *ne* occurred with no post-verbal negator were likewise not considered for analysis.[40] Primarily concerning Near-NSs and NSs, pleonastic *ne* without a post-verbal negator (e.g., *avant que je ne rentre* ('before I return')) were excluded, as well as fixed expressions such as *je ne sais quoi* ('I don't know').[41]

Generally, the presence of the consonantal /n/ segment provided sufficient evidence for the presence of *ne,* as many tokens contained /n/ in non-elision contexts without vocalic /ø/ or /ə/ (e.g. /ʒønsepa/, transcribed as *je n'sais pas*). In some cases, it was impossible to determine within reasonable certainty whether *ne* was retained or not. Often, this was due to the participant producing inaudible speech. Sometimes the participant's own volume was low at the point of the negation utterance, or there was unclear enunciation. In other instances, there was overlapping speech with the interlocutor, and a clear decision could not be made; I excluded these tokens from

[40] There were two examples of *ne* produced with verbs that allow for the semantic content of negation to be expressed solely with *ne* rather than with a post-verbal negator, one by speaker 2P (see (1a)) and one by bilingual interlocutor 'Th' (see (2a)). In both cases, omission of *ne* would have resulted in a situation where the intended meaning of verbal negation could not be assumed (see (1b and 2b)). Since variation with *ne* is not possible in the context of these negated utterances due to the lack of post-verbal negator, these two tokens were not included in the statistical analyses.

(1a) *Je n'oserais imaginer*.
'I wouldn't dare imagine.'

(1b) *J'oserais imaginer.*
'I would dare imagine.'
* 'I wouldn't dare imagine.'

(2a) *On ne cesse de faire des présentations orales.*
'We have not stopped doing oral presentations,' i.e., 'We've been doing oral presentations non-stop.'

(2b) *On cesse de faire des présentations orales.*
'We have stopped doing oral presentations.'
* 'We have not stopped doing oral presentations.'

[41] The fixed expression *je ne sais quoi* was produced five times in the corpus, by Near-NSs 2P (three occurrences) and 5P, and NS KeF. All occurrences were modeled on the syntactically similar *je ne sais pas* ('I don't know') rather than using the expression as a syntactic subject (e.g., *un certain je ne sais quoi* ('a certain je ne sais quoi')). It is possible that L2 French speakers could delete *ne* in the former example by analogy to *je sais pas*, but such an example (??*je sais quoi*) was not produced in this corpus.

the statistical analysis. Finally, it is noteworthy to mention that some utterances contained no clear indication of consonantal /n/ in my judgment, but in the context before where a /n/ could occur (e.g., /ʒøːsepa/, *je sais pas*), the speaker lengthened the vowel somewhat or produced a slightly nasalized vowel or, in the case of subject pronoun *on* followed by a consonant, a lengthened nasalized vowel (e.g., /ɔ̃ːsepa/, *on sait pas*). In these cases, it is impossible to clearly determine if the speaker intended to produce *ne*.[42] In his analysis of *ne*-retention, Donaldson (2017) established a rating system where he rated the presence or absence of *ne* for each token on a scale of 1 ("absolutely cannot tell") to 5 ("absolutely confident"); tokens rated as 1 or 2 ("fairly doubtful") were excluded. Though I did not adopt this kind of rating system, I had a similar threshold for inclusion. With the Express Scribe software, a quick and repeated review of each ambiguous utterance was possible by using one keystroke to rewind the audio by one second. After several reviews, if I could not judge the token as either retention or omission, I noted it but did not include it in the analysis. All judgments were my own.

This process resulted in over 5000 tokens for all speaker groups and interlocutors combined. In the following subsection, the results for SA learners and interlocutors will be examined first, to be followed by results for Near-NSs and their interlocutors.

4.4.1 SA Learners: Overall results

Table 4-4 summarizes variable *ne*-retention in verbal negation contexts for all SA learners and interlocutors. In all, 804 tokens of verbal negation were coded for the presence or absence of *ne.* An additional seven verbal negation contexts by SA learners and one context by near-native/NS

[42] One possibility for determining these more ambiguous cases would involve an acoustic analysis using more sophisticated tools to determine whether the nasalization occurred to such an extent that the utterance could be perceived as realizing *ne*. In this case, it would be necessary to quantitatively establish a sufficient (and consistent) threshold for the production of a segment that would be considered as a realization of the /n/ phoneme and therefore, a token of *ne*-retention. However, such fine-grained analysis is beyond the scope of this study.

interlocutors were determined to be inconclusive regarding the presence or absence of *ne*; these tokens (representing 1.0% of all verbal negation utterances) were not included in the following calculations.

Table 4-4. Ne-retention in SA learner group

Speaker ID	*ne* tokens	Total verbal negation contexts	% *ne*-retention
1S	19	39	48.7
2S	28	56	50.0
3S	61	76	80.3
4S	27	60	45.0
5S	42	65	64.6
6S	51	52	98.1
7S	50	51	98.0
8S	46	95	48.4
Overall L2 (SA learners)	**324**	**494**	**65.6**
L2 near-native interlocutor	**16**	**195**	**8.2**
L1 native interlocutor	**6**	**115**	**5.2**

There is much variation in *ne*-retention among SA learners, with two approaching categorical retention, despite the low retention of the native and near-native interlocutors. As one can easily observe based on the percentages, SA learners demonstrate significantly higher *ne*-retention than their interlocutors ($\chi^2(1) = 266$; $p < .0001$). By way of comparison with other studies, the 65.6% *ne*-retention in SA learners is situated squarely between the results from learners in Dewaele (2004a), 72.8%, and Sax (2003), 60.8%. Moreover, the similar retention rates of the near-native (8.2%) and native speaker (5.2%) interlocutors are much lower than those reported in Coveney (2002), 19%, and in Dewaele (2004a), 36.3%, but similar to overall NS interlocutor rates with the Near-NS groups in the current study (7.6%).

The *c*-test score measuring proficiency generally did not correlate with *ne*-retention rates in SA learners (see Figure 4-1). As a group, however, there was a broader correlation between the participants' language security index (refer to section 3.5.1.4 for details on how this index was

calculated) and their *ne*-retention rates, as Figure 4-2 shows. Note that in both figures, *ne* is expressed in terms of *deletion* (represented by the orange lines) for ease of comparison, with the expectation that higher *ne*-deletion would generally correspond to higher proficiency scores. The *c*-test proficiency scores and security index results are represented by the blue bars.

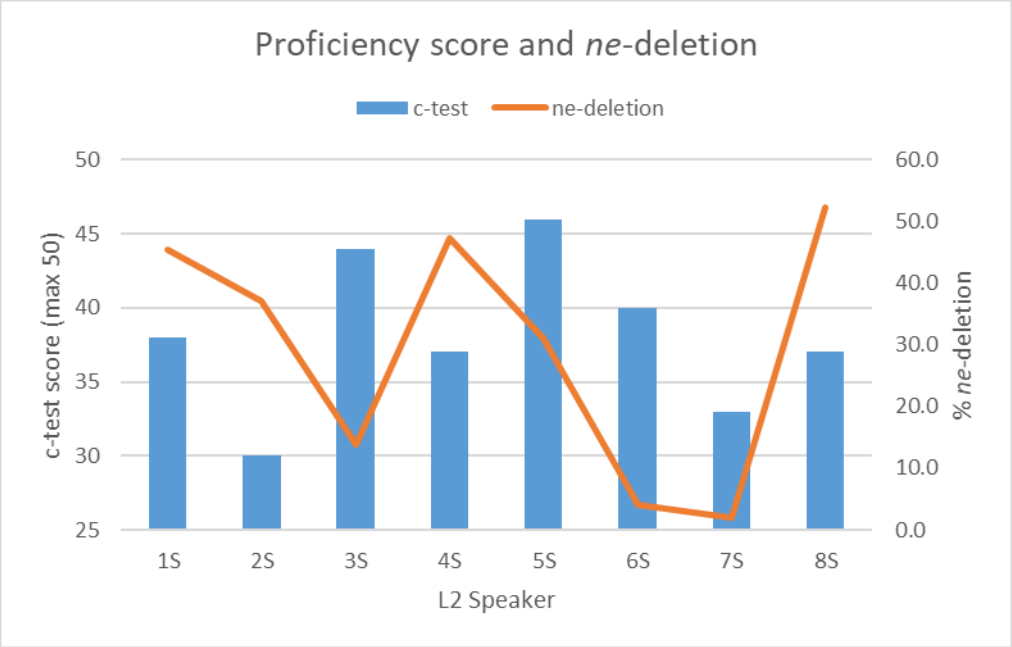

Figure 4-1. C-test proficiency score and ne-deletion in SA learners

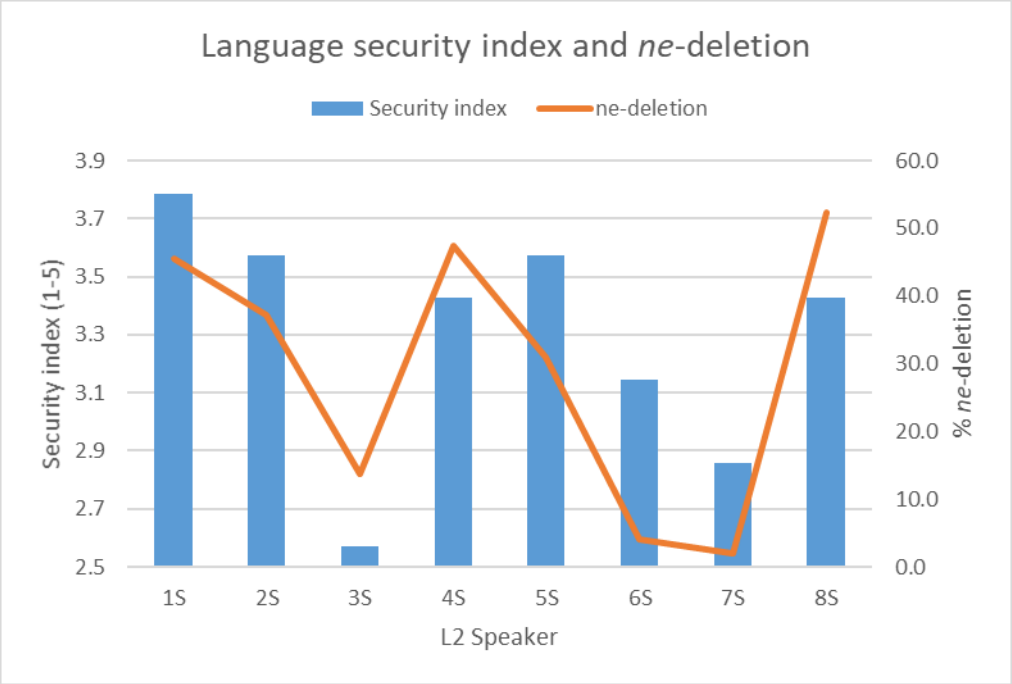

Figure 4-2. Language security index and ne-deletion in SA learners

If the *c*-test score directly correlated with *ne*-deletion, we would expect higher *ne*-deletion for speakers 3S, 5S, 6S, and 7S, as well as lower *ne*-deletion for the speaker with the lowest *c*-test score (2S). For the language security index, one expected finding is that the two speakers with the shortest length of time studying French (3S at three years of French, 7S at 1.5 years of French) not only have the lowest language security indices, but also two of the three lowest *ne*-deletion rates. From this observation, one can suggest that, though these speakers were able to perform comparatively well on a written proficiency task (given their length of time studying French), their lower language security may better explain their lack of sensitivity to sociolinguistic variation. However, note that we would still expect a lower security index for speaker 6S and, possibly, a higher index for speaker 8S, based on their *ne*-deletion. Nevertheless, the general correlation between the security index and *ne*-deletion may indicate that as learners become more comfortable

in their use of the L2 (but not necessarily more proficient on written tasks), they omit *ne* more in line with native speaker tendencies.

4.4.2 SA Learners: *Ne*-retention by interlocutor type

I now turn to the *ne*-retention rates obtained from the two interlocutors for the SA learners. Unsurprisingly, these rates are much lower than those found in the SA learners. Table 4-5 provides these figures again for the near-native (AmE) and native (SoF) interlocutor.

Table 4-5. Ne-retention in the native and near-native interlocutor (with SA learner group)

Speaker ID	***ne* tokens**	**Total verbal negation contexts**	**% *ne*-retention**
L2 near-native interlocutor (AmE)	16	195	8.2
L1 native interlocutor (SoF)	6	115	5.2
Total	**22**	**310**	**7.1**

The overall rates for each speaker fall within the ranges for recent studies on L1 French and L2 (near-native) French, suggesting that these speakers are representative, with respect to *ne*-retention, of the kind of informal speech patterns that the SA learners may encounter elsewhere in the target language community. The difference in *ne*-retention rates between these two speakers is not significant ($\chi^2(1) = 0.979$; $p = .322$).

Concerning SA learners' *ne*-retention according to the language status of each interlocutor, Table 4-6 reports the results of SA learners, with Dewaele's (2004a) results by way of comparison.

Table 4-6. Ne-retention rates in SA learners divided by interlocutor L1 status

	Overall results (all speakers)		**L1 and L2 results divided by L1 status**			**SA learners: results across interlocutor type**		
	n	Total negation contexts	French status	*ne* / total neg	% *ne*	Interlocutor type	*ne* / total neg	% *ne*
Dewaele (2004a)	73	991	L1	N/A	36.3	L1 French	N/A	53.5
			L2	N/A	72.8	L2 French	N/A	75.5
SA learners[43]	10	804	L1	6/115	5.2	L1 French	90/162	55.6
			L2	324/494	65.6	L2 French	234/332	70.5

Recall that, overall, SA learners had strikingly similar *ne*-retention rates with Dewaele's learners (72.8% versus 65.6%; an apparent non-significant difference based on Dewaele's number of tokens). Furthermore, it is noteworthy that the NS interlocutor's retention rates in the current study are much lower than Dewaele's NS interlocutors. This may be due to a number of factors, including the demographics of the L1 speakers in Dewaele's study as well as Dewaele's inclusion of both formal and informal topics. In the current study, the participants were provided with suggestions of topics, none of which would be included in the category of "serious" topics (see Donaldson, 2017: 142); as a result, the vast majority of negation tokens (over 90%) were uttered in the context of likely neutral or informal topics. Moreover, *ne*-retention in SA learners when speaking to a native (55.6%) versus a non-native (70.5%, including conversations with the near-native and with other SA learners) patterns in the same direction as in Dewaele's study (see the last column in Table 4-6); this difference by interlocutor native language was significant in SA learners ($\chi^2(1) = 10.7$; $p = .001$).

Given the characteristics of the interlocutors for the SA learners, however, a more nuanced analysis than "native" versus "non-native" can reveal more about how the SA learners interact

[43] Totals reported in this column include the eight SA learners, the NS interlocutor, and the near-native interlocutor.

with interlocutors of different proficiency levels. Table 4-7 breaks down *ne*-retention results for SA learners across each of the three types of interlocutors.

Table 4-7. SA learners: Results across interlocutor type

Interlocutor type	***ne* / total negation tokens**	**% *ne*-retention**
L1 French: NS	90/162	55.6
L2 French: near-native	101/169	59.8
L2 French: SA learner	133/163	81.6

As the above table shows, overall, SA learners essentially treated the NS interlocutor and the near-native interlocutor no differently with regard to *ne*-retention (a highly non-significant difference: ($\chi^2(1) = 0.600$; $p = .439$). However, with their fellow SA learners as interlocutors, they produced significantly higher *ne*-retention ($\chi^2(1) = 27.6$; $p < .0001$) compared to the rates with the NS/near-native (grouped together). The simple explanation is that these learners considered the NS and near-native as some sort of model speakers of French, integrated into the Francophone community, whether or not these learners could recognize the slight differences (and possible non-targetlike deviations) in the near-native's speech, and despite the fact that the near-native made no explicit effort (nor was instructed) to conceal her American identity. Moreover, these SA learners had had almost no interaction with the NS or near-native beforehand; yet, they were more informal in their use of *ne* with these interlocutors. On the other hand, despite knowing their SA learner interlocutors beforehand, they were much more conservative in *ne*-retention. Again, a *prima facie* explanation is that, rather than "relax" their use of *ne* with familiar interlocutors, they adopted a more formal, more classroom-like use of French, at least with regard to *ne*-retention. These discussions will be revisited in Chapter 6.

These trends largely hold across individual speaker variation in SA learners as well. Figure 4-3 shows *ne*-retention rates across interlocutor type for each of these eight learners.

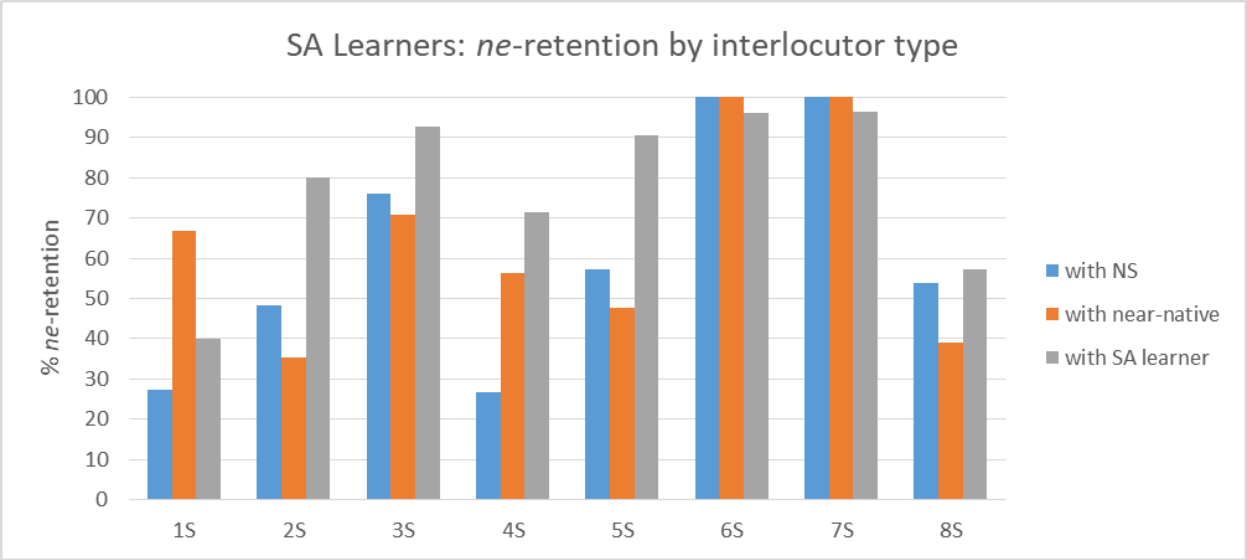

Figure 4-3. Ne-retention across interlocutor type for SA learners

As this figure indicates, in the majority of cases (five speakers of eight), *ne*-retention is highest in conversation with other SA learners. The three exceptions include Speaker 1S, whose highest *ne*-retention was with the near-native interlocutor, and Speakers 6S and 7S, who show nearly categorical *ne*-retention but one token each of *ne*-deletion with their SA interlocutors. As mentioned in section 3.6.3, the single cases of *ne*-deletion for these speakers warrant further discussion. Speaker 6S's lone deletion token may be considered marginal, with a slight pause (between one-third and one-half second in duration) between the verb and the negator: *mais c'est…pas le meilleur pour moi* ('but it's…not the best for me'). In cases where there is a pause in searching for words, native and non-native speakers alike may produce a verbal negator (in this case, *pas*) that was not yet anticipated at the moment of uttering the verb, and so utterance of pre-verbal *ne* would not be possible (or expected). Generally, when anticipating a negator after such pauses, other speakers simply repeat the subject-verb structure (e.g., *mais c'est…c'est pas le meilleur*), leading to a new opportunity for *ne*-retention; speakers of lower proficiency may either

deem such repair unnecessary or lack the fluency to repeat the entire structure. In other instances of hesitation with similar verbal structures, this pause was more apparent (closer to a full second), and such tokens (seven overall) were not included in the results. As for Speaker 7S, his lone token of *ne*-deletion was an immediate repetition of his interlocutor's identical utterance (*C'est pas cher*, 'It's not expensive'). While re-insertion of *ne* in a repetition of the original *ne*-deleted utterance is certainly possible, the fact that this was an immediate repetition, and the only possible example of *ne*-deletion in this speaker, also renders this token marginal.

Concerning Speaker 1S, who has higher retention with the near-native than with either of his other two interlocutors, one possible explanation for his behavior may be found in a disclosure during the debriefing; when asked which conversation seemed to be the most difficult in terms of ease of expression in French, he was the only SA learner to choose the conversation with the near-native. This difficulty may have resulted in the adoption of a more formal style, leading to higher *ne*-retention.

Furthermore, note that each type of conversation is likely to yield greater individual variation when fewer tokens are produced. Across all learners, the total tokens for each conversation ranged from nine to 41. Table 4-8 provides the distribution of these figures.

Table 4-8. Individual results of ne-retention for each type of SA learner conversation

Speaker	with NS		with near-native		with SA learner		Total	
	ne/total	% *ne*-retention	*ne*/total	% *ne*-retention	*ne*/total	% *ne*-retention	*ne*/total	% *ne*-retention
1S	3/11	27.3	12/18	66.7	4/10	40.0	19/39	48.7
2S	14/29	48.3	6/17	35.3	8/10	80.0	28/56	50.0
3S	19/25	76.0	17/24	70.8	25/27	92.6	61/76	80.3
4S	8/30	26.7	9/16	56.3	10/14	71.4	27/60	45.0
5S	12/21	57.1	11/23	47.8	19/21	90.5	42/65	64.6
6S	9/9	100.0	17/17	100.0	24/25	96.0	50/51	98.0
7S	11/11	100.0	13/13	100.0	27/28	96.4	51/52	98.1
8S	14/26	53.8	16/41	39.0	16/28	57.1	46/95	48.4
Total	90/162	55.6	101/169	59.8	133/163	81.6	324/494	65.6

As this table shows, Speaker 1S produced the fewest negation tokens of all SA learners. For the remaining speakers showing variation, higher overall numbers of negation tokens are consistent with higher *ne*-retention with SA interlocutors. Other considerations aside, these results suggest an influence of interlocutor effect detectable beyond variation in individual speaker patterns.

4.4.3 Near-NSs: Overall results

For the Near-NS groups and their interlocutors, conversations retained for analysis (that is, excluding speakers not meeting near-native criteria) included 4,305 tokens of verbal negation that I coded for the presence or absence of *ne*, with 570 instances of *ne*-retention (for a global rate of 13.2% retention combining Near-NSs and interlocutors). An additional 19 verbal negation contexts by Near-NSs and 15 contexts by NSs were determined to be inconclusive regarding the presence or absence of *ne*; these tokens (representing 0.8% of all verbal negation utterances) were not included in the following calculations. Table 4-9 gives the overall results of negation usage for Near-NSs and interlocutors in Pau, followed by Table 4-10 for Near-NSs and interlocutors in Lille.

Table 4-9. Ne-retention in Near-NSs in Pau

Speaker ID: Pau	***ne* tokens**	**Total verbal negation contexts**	**% *ne*-retention**
1P	40	100	40.0
2P	57	168	33.9
3P	10	103	9.7
4P	15	162	9.3
5P	40	178	22.5
6P	34	67	50.7
7P	29	141	20.6
8P	2	118	1.7
9P	23	98	23.5
10P	2	89	2.2
Overall Near-NSs	**252**	**1224**	**20.6**
Bilingual interlocutors (English identity)	36	606	5.9
Bilingual interlocutors (French identity)	47	683	6.9
Overall bilingual interlocutors	**83**	**1289**	**6.4**

Table 4-10. Ne-retention in Near-NSs in Lille

Speaker ID: Lille	***ne* tokens**	**Total verbal negation contexts**	**% *ne*-retention**
1L	4	65	6.2
2L	0	95	0.0
3L	23	100	23.0
4L	43	177	24.3
5L	1	69	1.4
6L	27	123	22.0
7L	11	64	17.2
8L	6	73	8.2
9L	14	71	19.7
Overall Near-NSs	**129**	**837**	**15.4**
L2 near-native interlocutors	41	321	12.8
L1 native interlocutors	65	634	10.3
Overall interlocutors	**106**	**955**	**11.1**

Across both research sites, overall *ne*-retention in Near-NSs is 18.5%, with 381 instances of *ne* in 2,061 variable verbal negation contexts. As expected, Near-NSs produce much lower *ne*-retention overall as compared with the SA learner group (65.6% versus 18.5%). In Pau, Near-NSs had significantly higher *ne*-retention than their bilingual interlocutors overall ($\chi^2(1) = 109$; $p < .0001$). In Lille, the difference was somewhat less pronounced, though this 4.3% difference between Near-NSs and their interlocutors overall was still highly significant ($\chi^2(1) = 7.28$; $p = .007$).

As for comparisons with other observations of near-native speakers, Near-NSs in Pau had similar *ne*-retention rates (20.6%) as in Donaldson's near-native speakers (22.4%); this difference was not significant ($\chi^2(1) = 1.30$; $p = .255$). Near-NSs in Lille had significantly lower *ne*-retention rates (15.4%) compared with speakers in Pau ($\chi^2(1) = 8.84$; $p = .003$) as well as compared with Donaldson's near-native speakers ($\chi^2(1) = 14.6$; $p < .0001$). The range of individual speaker frequencies across both sites (0% to 50.7%) is larger than in Donaldson's Near-NSs (4.8% to 38.0%), though in Lille the range is smaller (0% to 24.3%). Regarding the order of conversations, Near-NSs in Pau had 18.8% *ne*-retention in the first conversation compared with 22.5% retention in the second conversation; this difference is not significant ($\chi^2(1) = 2.65$; $p = .104$). In Lille, the difference between the first (18.5%) and the second (12.9%) conversation is significant ($\chi^2(1) = 5.10$; $p = .024$), but only when excluding near-native interlocutor SaE's conversation with the NS (otherwise non-significant: $\chi^2(1) = 2.7$; $p = .096$).[44] Thus, it is possible that lower *ne*-retention in Lille may be influenced by a combination of interlocutor L1 and order of conversations.[45]

4.4.3.1 Near-NSs: Interlocutor results

The Near-NS interlocutors' overall *ne*-retention rates (8.6%, 198/2304) are much lower than in Coveney (2002; 19%) but only slightly lower than Donaldson's (all-native-speaker) interlocutors (11.1%). Native speaker interlocutors, which includes all interlocutors in Pau and the native interlocutors in Lille, produce *ne* at a rate of 7.7% (148/1923). This rate is slightly lower

[44] Recall that interlocutor SaE was initially recruited as a Near-NS participant (recording a conversation with the NS interlocutor CaF) and subsequently served as primary near-native interlocutor when the originally recruited near-native interlocutor declined to participate due to scheduling conflicts.

[45] In Lille, the combination with the largest deviance from the overall Near-NS *ne*-retention average is with NS interlocutors in the first conversation (20.0%). Since this combination represents only three conversations, the effect of order of conversation found in Lille may be explained by small sample sizes.

than that of Donaldson's native speakers, and with a large number of tokens, produces a significant difference between native speaker groups across both studies ($\chi^2(1) = 8.67$; $p = .003$).

The similarities in interlocutor rates at each site are striking. In Pau, the adoption of an English versus French identity did not produce significant differences in *ne*-retention in the bilingual interlocutors as a whole (5.9% *ne*-retention overall with an English identity compared with 6.9% *ne*-retention overall with a French identity, $\chi^2(1) = 0.472$; $p = .492$), though one of the bilinguals had large variation across identities (albeit with a small sample size). In Lille, the difference between L1 French interlocutors and L2 French interlocutors (10.3% versus 12.8%, respectively), as a whole, was statistically insignificant ($\chi^2(1) = 1.37$; $p = .242$).

Due to the specific nature of the interlocutors recruited for this study, including the bilingual identities in Pau and the necessity of recruiting additional interlocutors due to scheduling conflicts, some important details emerge when examining interlocutor *ne*-retention behavior in more detail. First, individual results for all three interlocutors in Pau, separated by English/French identity, are indicated in Table 4-11.

Table 4-11. Individual ne-retention rates for bilingual interlocutors in Pau; n = number of conversations for each identity

Speaker ID	***ne* tokens**	**Total verbal negation contexts**	**% *ne*-retention**
ChE (English identity; *n* = 4)	12	308	3.9
ChF (French identity; *n* = 5)	11	406	2.7
Ch (all)	**23**	**714**	**3.2**
FrE (English identity; *n* = 4)	5	205	2.4
FrF (French identity; *n* = 4)	6	230	2.6
Fr (all)	**11**	**435**	**2.5**
ThE (English identity; *n* = 2)	19	93	20.4
ThF (French identity; *n* = 1)	30	47	63.8
Th (all)	**49**	**140**	**35.0**
All English identity	36	606	5.9
All French identity	47	683	6.9
Overall bilingual interlocutors	**83**	**1289**	**6.4**

As the table shows, Ch and Fr had low overall *ne*-retention rates, concordant with other recent studies of native speakers, and their *ne*-retention rates under both guises were near-identical (Ch: ($\chi^2(1) = 0.791$; $p = .374$); Fr: ($\chi^2(1) = 0.013$; $p = .910$)). In contrast, Th had significantly higher *ne*-retention overall. Th's results here include a conversation at 63.8% retention with speaker 1P, which strongly skews his overall average due to the lower token counts produced from having participated in only three conversations, and this single conversation also skews the comparison across Th's English and French guises, in which there was a significant difference in *ne*-retention across both identities ($\chi^2(1) = 25.8$; $p < .0001$). This sample size limits the ability to gauge how much (or whether) *ne*-retention was consciously or unconsciously influenced by his adoption of one or the other identity. If, however, we look at *all* of Th's conversations, including those with speakers who did not meet the criteria for near-native status (identified as 11P and 12P in Table 4-12), we see that Th's overall *ne*-retention (24.1%) is much lower than that reported in Table 4-11. In these two additional conversations (one under each identity), Th produced 8.6% and 9.1% retention, rates that may more closely reflect Th's use of *ne* in informal contexts in general, as compared with his 35.0% retention rate in the three conversations in Table 4-11,[46] and reducing the likelihood that the adoption of a French identity alone explains his high *ne*-retention.

[46] The *ne*-retention rates for the bilingual interlocutors in the remaining conversations with speakers who did not meet Near-NS criteria are as follows: Ch (French identity) with 11P: 2.1% (1/47); Fr (English identity) with 12P: 1.8% (1/57).

Table 4-12. Individual ne-retention rates for bilingual 'Th' for each conversation

Near-NS interlocutor and 'Th' identity	***ne* tokens**	**Total verbal negation contexts**	**% *ne*-retention**
1P (French)	30	47	63.8
3P (English)	15	69	21.7
4P (English)	4	24	16.7
11P (English)	3	35	8.6
12P (French)	6	66	9.1
All English identity	22	128	17.2
All French identity	36	113	31.9
Totals	**58**	**241**	**24.1**

Moreover, the order in which Th's conversations were recorded does not reflect a pattern in which Th had high *ne*-retention in his first conversation and then lower rates in subsequent conversations; Th's first two conversations were with speakers 11P and 12P, with whom Th had 8% *ne*-retention overall. Rather, a likely possibility is that 1P's biographical characteristics (as a much older female) strongly influenced Th's perception of formality for the conversation, as the dyad was mismatched for gender and, of all Th's conversations, was most mismatched for age. The remaining conversations were matched for gender (12P) or had smaller age gaps.

Even with these considerations in mind, Th still produced higher *ne*-retention than his bilingual counterparts in the remaining conversations. A potential explanation is that after the recording sessions, Th expressed to me his concerns about properly fulfilling his role as conversation partner, indicating that he may have more highly monitored his speech, at least in certain conversations, whereas Ch and Fr indicated no such difficulty.

Indeed, *ne*-retention rates for Ch and Fr are relatively stable across all conversations, as Table 4-13 and Table 4-14 show:

Table 4-13. Individual ne-retention rates for bilingual Ch for each conversation

Near-NS interlocutor and Ch identity	*ne* tokens	Total verbal negation contexts	% *ne*-retention
1P (English)	1	56	1.8
2P (French)	4	60	6.7
4P (French)	1	70	1.4
5P (English)	2	68	2.9
6P (French)	2	78	2.6
7P (English)	5	83	6.0
8P (English)	4	101	4.0
9P (French)	2	89	2.2
10P (French)	2	109	1.8
Totals	**23**	**714**	**3.2**

Table 4-14. Individual ne-retention rates for bilingual Fr for each conversation

Near-NS interlocutor and Fr identity	*ne* tokens	Total verbal negation contexts	% *ne*-retention
2P (English)	4	28	14.3
3P (French)	1	82	1.2
5P (French)	3	61	4.9
6P (English)	0	83	0.0
7P (French)	1	39	2.6
8P (French)	1	48	2.1
9P (English)	0	44	0.0
10P (English)	1	50	2.0
Totals	**11**	**435**	**2.5**

As these tables indicate, the highest retention for both of these bilinguals is in conversation with speaker 2P. Potential explanations may involve a higher level of monitoring on the part of the bilinguals, likely due to 2P's status as the most nativelike Near-NS, discussion of her career as professional translator, and the predominance of topics concerning "correct" spoken and written forms of language in both conversations.

Regarding interlocutors in Lille, recall that these speakers were recruited based on their status as either near-native or native French speakers, and that due to logistical limitations, one additional Near-NS and one additional NS were recruited to serve as conversation partners. Table 4-15 gives the overall *ne*-retention rates for all interlocutors.

Table 4-15. Individual ne-retention rates for native and near-native interlocutors in Lille

Speaker ID	*ne* tokens	Total verbal negation contexts	% *ne*-retention
CaF (NS; *n* = 7)[47]	58	532	10.9
KeF (NS; *n* = 2)	7	102	6.9
Overall NS interlocutors	**65**	**634**	**10.3**
SaE (near-native; *n* = 6)[48]	32	229	14.0
JeE (near-native; *n* = 3)	9	92	9.8
Overall near-native interlocutors	**41**	**321**	**12.8**
Overall interlocutors	**105**	**960**	**10.9**

Though the number of tokens being compared is smaller (and unevenly distributed), there is no statistically significant difference in *ne*-retention rates between each of the native ($\chi^2(1) = 1.52$; $p = .218$) and near-native ($\chi^2(1) = 1.03$; $p = .309$) interlocutors; moreover, as one may expect, the 2.5% percent difference between the overall NS interlocutor and near-native interlocutor rates is not significant ($\chi^2(1) = 1.37$; $p = .242$).

4.4.3.2 Near-NSs: Correlations with the Acceptability Judgment Task

Since *c*-test scores for Near-NSs were at ceiling for most speakers, a comparison of *c*-test scores with *ne*-retention would be of little value compared with the SA learners. With Near-NSs, comparisons can be made with *ne*-retention rates and the language security index based on the language background questionnaire, as well as the Acceptability Judgment Task (for those Near-NSs who completed this task). For the AJT, for each speaker I can compare the average deviation from NSs in previous studies (as reported in section 3.5.2.6) and identify whether there is a correlation with *ne*-retention. Recall that for each of the 76 AJT items, I took the average rating of the AJT results for NSs reported in Birdsong (1992) and Donaldson (2008) and computed the

[47] CaF produced 15.0% (9/60) *ne*-retention in the conversation with SaE that was not included in the analysis. If included, the overall percentages for both CaF individually and for the two native interlocutors combined would increase by 0.4% each. This does not affect the significance of the statistical tests.

[48] SaE produced 4.7% (2/43) *ne*-retention in the conversation with CaF; these results are not included here.

difference for each Near-NS and interlocutor (including NSs) who completed the task. I then averaged each difference to determine how much each speaker deviated from Birdsong's and Donaldson's NSs. The results from this calculation are reproduced in Table 4-16 along with each speaker's overall *ne*-retention rate.

Table 4-16. Average deviation from NSs on AJT compared with ne-retention

Speaker ID	**AJT: Avg. deviation from NSs**	**% *ne*-retention (overall)**
1P	0.711	40.0
2P	0.383	33.9
3P	0.706	9.7
4P	1.077	9.3
5P	0.760	22.5
6P	0.872	50.7
8P	0.849	1.7
9P	0.851	23.5
10P	0.889	2.2
1L	0.743	6.2
2L	0.852	0.0
3L	0.885	23.2
6L	0.847	22.0
7L	0.845	17.2
8L	0.746	8.2
9L	1.119	19.7
SaE	0.868	14.0
JeE	0.684	9.8
CaF	0.422	10.8
KeF	0.619	6.9
Ch	0.706	3.2
Fr	0.608	2.5

Based on these comparisons, there is no clear correlation between nativelike AJT results and low *ne*-retention. To represent this lack of correlation more visually, Figure 4-4 shows AJT average deviations (blue bars) along with overall *ne*-retention percentages (orange line) for each speaker who completed the AJT.

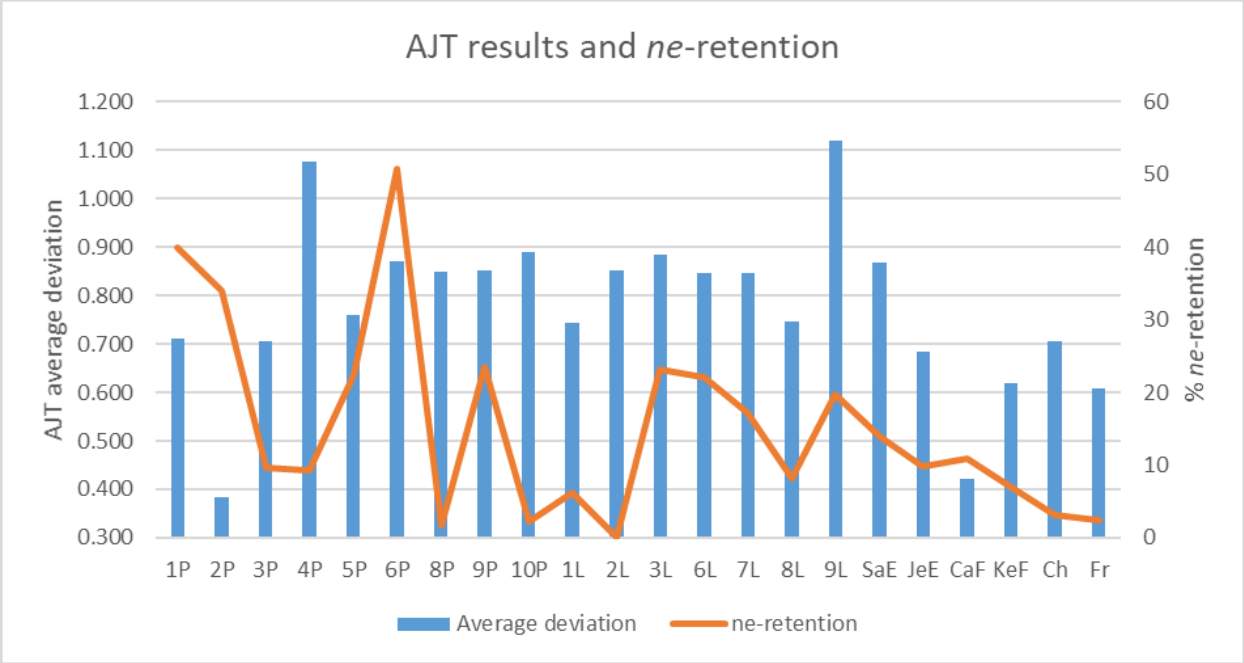

Figure 4-4. AJT results and ne-retention.

From this figure, two initial observations can be made. First, there is relative homogeneity in the AJT average deviations, with two speakers (4P and 9L) who are more than one standard deviation above the mean (SD = .174) and three speakers (2P, CaF, and Fr) more than one standard deviation below the mean. Second, note that the highest *ne*-retention rates (in 1P, 2P, and 6P) are not predictive of less nativelike AJT results, and the speakers with the least nativelike AJT results (4P and 9L) actually have *ne*-retention rates close to or below the Near-NS average.

4.4.3.3 Near-NSs: Correlations with the language security index

As for the language security index and its relationship with *ne*-retention in Near-NSs, the results are also somewhat inconclusive. Figure 4-5 and Figure 4-6 for Pau and Lille, respectively, plot speakers' language security indices and *ne*-deletion rates. Note that these figures represent *deletion* in order to graphically visualize a possible positive correlation between language security and informal use of *ne*; note also that the scale for *ne*-deletion in these figures begins at 40%.

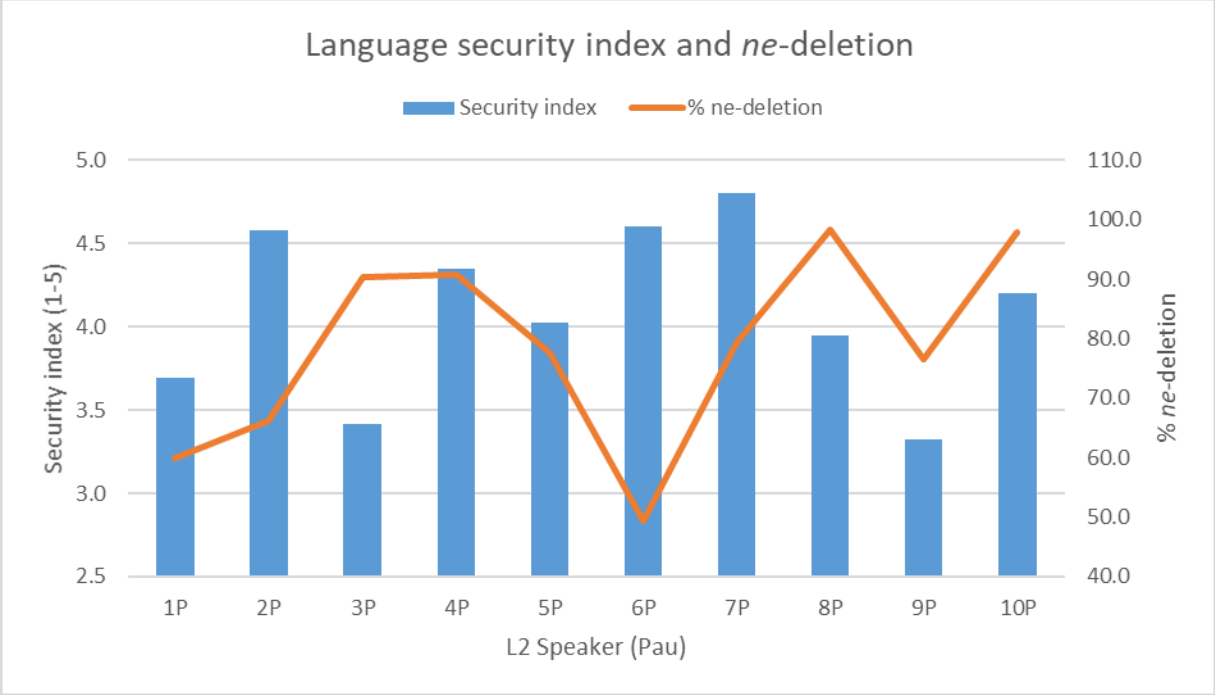

Figure 4-5. *Language security index and ne-deletion for Near-NSs in Pau*

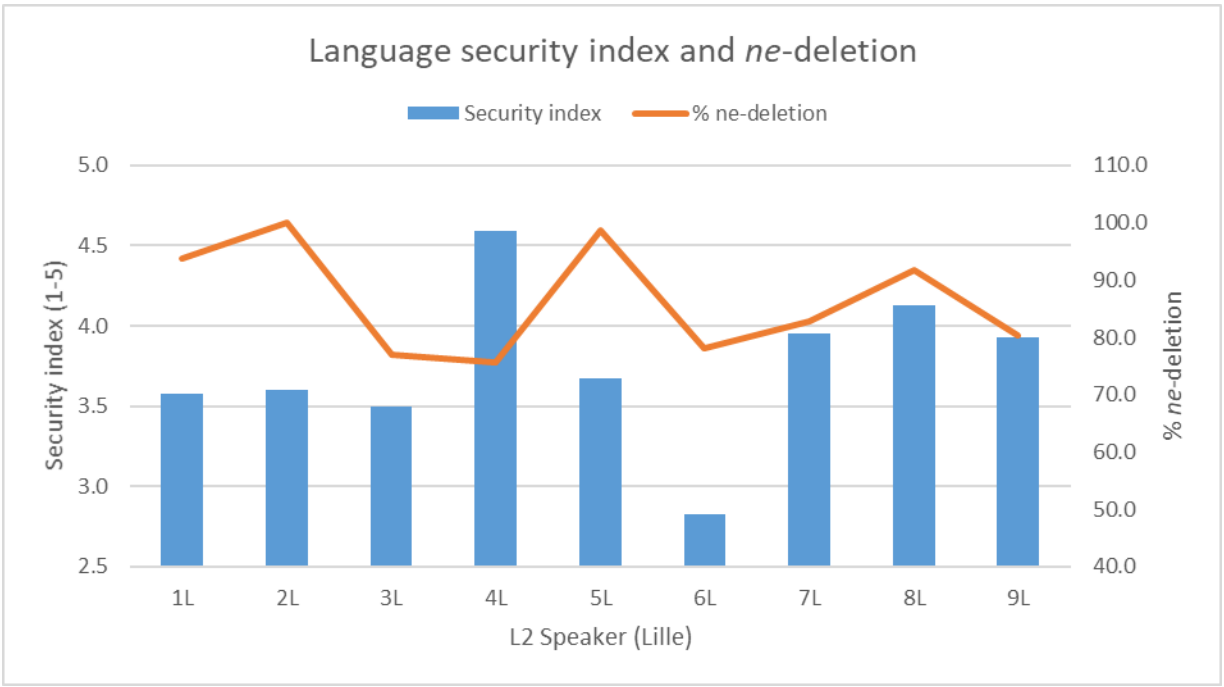

Figure 4-6. *Language security index and ne-deletion for Near-NSs in Lille*

It appears that language security has very little to do with *ne*-deletion in for Near-NSs in Pau. In Lille, there does seem to be a slight correlation for several speakers, despite the fact that *ne*-deletion ranges are much smaller in Near-NSs in Lille than in Pau. The exceptions are speaker 4L, for whom one would expect higher *ne*-deletion based on his language security index, and speaker 6L, for whom one would expect much lower *ne*-deletion.

These results may be partly explained by other factors. Speakers 2P and 4L were the most proficient speakers in Pau and Lille, respectively, based on my informal observations as well as by an error analysis (I detected no grammatical errors in 2P's conversations and one error in 4L's conversations) and by their high language security indices (4L had, by far, the highest language security index in Lille). It is possible that *ne*-retention is influenced by relative effects of age or profession (as observed in the pilot study), which may result in more conservative sociolinguistic variation; indeed, the three highest *ne*-retention rates in Pau (1P, 2P, 6P) belong to three of the four oldest speakers, and for speakers 2P and 6P, their professions may play a role as well (professional translator and former French teacher, respectively), which may also account for their high language security. Age did not affect *ne*-retention in the same way for the oldest Lille speaker (5L), while 4L's professional training as a linguist may have played a role in his conservative use of *ne* (conversely, 4L's linguistics training could render him *less* concerned about prescriptive norms, though this behavior would not be accounted for by his relatively high *ne*-retention). Finally, it is possible that these highly proficient speakers are simply motivated by non-native pragmatic conservatism to retain *ne*, as Donaldson (2017) observed for some of his near-native speakers (cf. section 2.9). As for speaker 6L, one may expect higher *ne*-retention based on his low language security score; however, his status as a younger speaker with liberal use of informal vocabulary

(as seen in section 3.6.2) may partly account for his comparatively low *ne*-retention, despite a higher rate of grammatical errors that is consistent with his self-reported proficiency.

4.4.4 Near-NSs: *Ne*-retention by interlocutor type

I now turn from discussion of overall *ne*-retention rates in Near-NSs to results obtained in each type of conversation. When analyzing *ne*-retention by interlocutor type, a comparison is warranted between Near-NSs in the current study and studies targeting learners at similar near-native levels, such as Donaldson (2017). Table 4-17 provides the results of *ne*-retention in Near-NSs in the current study, separated by data collection site and divided by interlocutor type (see third column). Results obtained by Donaldson as well as by Dewaele (2004a) are reported for comparison purposes.

Table 4-17. Ne-retention rates in Near-NSs divided by interlocutor L1 status

	Overall results (all speakers)		**L1 and L2 results divided by L1 status**			**L2 speakers: results across interlocutor type**		
	n	Total negation contexts	French status	*ne* / total neg	% *ne*	Interlocutor type	*ne* / total neg	% *ne*
Dewaele (2004a)	73	991	L1	N/A	36.3	L1 French	N/A	53.5
			L2	N/A	72.8	L2 French	N/A	75.5
Donaldson (2017)	20	1877	L1	108/985	11.0	L1 French	202/892	22.4
			L2	202/892	22.4			
Near-NS: Pau	13	2513	L1[49]	83/1288	6.4	L1 French ID	139/635	21.9
			L2	252/1224	20.6	L1 English ID	113/589	19.2
Near-NS: Lille	11	1792	L1	65/634	10.3	L1 French	79/501	15.8
			L2[50]	170/1158	14.7	L2 French	50/336	14.9

In both sites in the current study, there were non-significant differences in *ne*-retention rates for Near-NSs across interlocutors of different L1 identities or native language status. Near-NSs in Pau

[49] Includes all results for the bilingual interlocutors, regardless of assumed identity.
[50] Includes Near-NSs and near-native interlocutors.

produced slightly higher rates of *ne*-retention in conversation with bilinguals adopting an L1 French identity (21.1%) compared with an L1 English identity (19.4%); this difference of 2.7% was not significant ($\chi^2(1) = 1.37$; $p = .242$). Near-NSs in Lille produced even more similar *ne*-retention percentages across both types of interlocutors: 15.8% with a NS versus 14.9% with a near-native, a non-significant difference ($\chi^2(1) = 0.122$; $p = .727$).

Focusing on differences for individual Near-NSs, Figure 4-7 and Figure 4-8 chart *ne*-retention rates across interlocutor type for all Near-NSs in Pau and Lille.

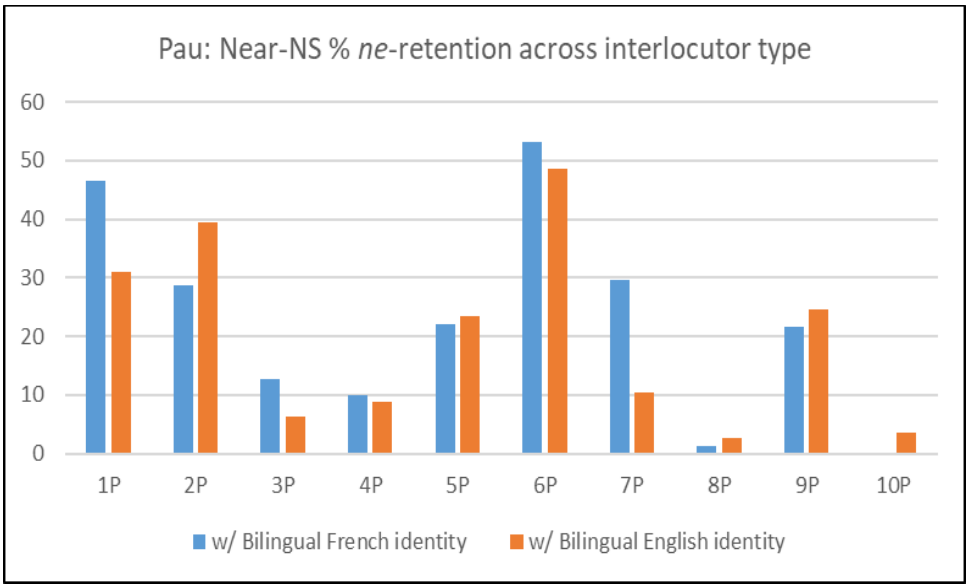

Figure 4-7. Ne-retention across interlocutor type for Near-NSs in Pau

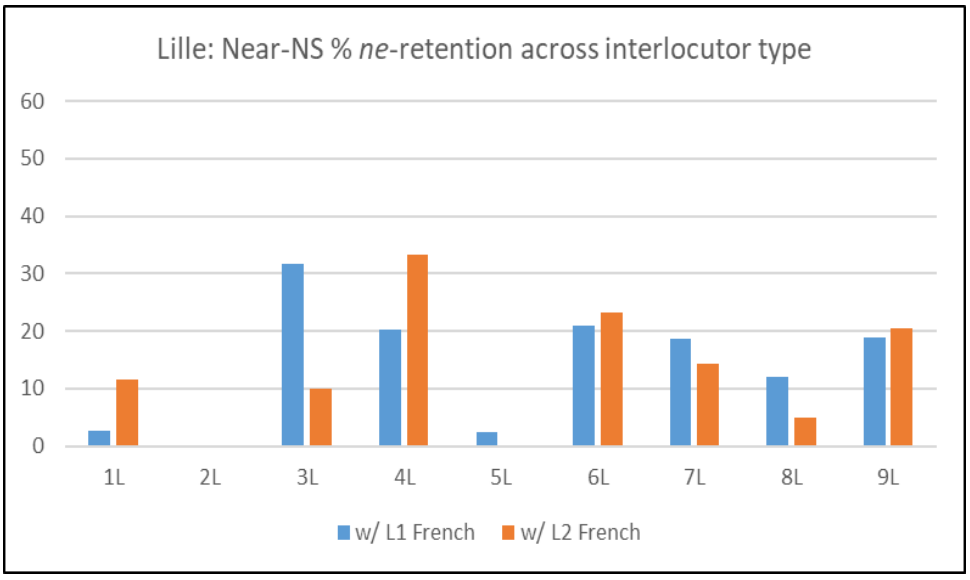

Figure 4-8. Ne-retention across interlocutor type for Near-NSs in Lille

As these figures illustrate, there is no identifiable trend toward preference of *ne*-retention in one interlocutor type. In Pau, five Near-NSs have higher *ne*-retention with English-identity bilinguals and five have higher *ne*-retention with French-identity bilinguals. In Lille, three Near-NSs have higher *ne*-retention with L1 French interlocutors, and five have higher *ne*-retention with L2 French interlocutors (with one speaker, 2L, producing no *ne*-retention across both conversations). As Table 4-18 (Pau) and Table 4-19 (Lille) show, the number of tokens per individual conversation is also generally higher than in the SA learner group, which is expected given the longer recording time and the more advanced fluency of Near-NSs.

Table 4-18. Near-NS individual results of ne-retention for each conversation (Pau)

Speaker	with Bilingual French identity		with Bilingual English identity		Total		Difference in % *ne*-retention
	ne/total	% *ne*-retention	*ne*/total	% *ne*-retention	*ne*/total	% *ne*-retention	
1P	27/58	46.6	13/42	31.0	40/100	40.0	+ 15.6
2P	25/87	28.7	32/81	39.5	57/168	33.9	- 10.8
3P	7/55	12.7	3/48	6.3	10/103	9.7	+ 6.5
4P	6/60	10.0	9/102	8.8	15/162	9.3	+ 1.2
5P	26/118	22.0	14/60	23.3	40/178	22.5	- 1.3
6P	17/32	53.1	17/35	48.6	34/67	50.7	+ 4.6
7P	22/74	29.7	7/67	10.4	29/141	20.6	+ 19.3
8P	1/81	1.2	1/37	2.7	2/118	1.7	- 1.5
9P	8/37	21.6	15/61	24.6	23/98	23.5	- 3.0
10P	0/33	0.0	2/56	3.6	2/89	2.2	- 3.6
Total	139/635	21.9	113/589	19.2	252/1224	20.6	+ 2.7

Table 4-19. Near-NS individual results of ne-retention for each conversation (Lille)

Speaker	with NS		with near-native		Total		Difference in % *ne*-retention
	ne/total	% *ne*-retention	*ne*/total	% *ne*-retention	*ne*/total	% *ne*-retention	
1L	1/39	2.6	3/26	11.5	4/65	6.2	- 9.0
2L	0/56	0.0	0/39	0.0	0/95	0.0	0.0
3L	19/60	31.7	4/40	10.0	23/100	23.0	+ 21.7
4L	25/123	20.3	18/54	33.3	43/177	24.3	- 13.0
5L	1/43	2.3	0/26	0.0	1/69	1.4	+ 2.3
6L	14/67	20.9	13/56	23.2	27/123	22.0	- 2.3
7L	8/43	18.6	3/21	14.3	11/64	17.2	+ 4.3
8L	4/33	12.1	2/40	5.0	6/73	8.2	+ 7.1
9L	7/37	18.9	7/34	20.6	14/71	19.7	- 1.7
Total	79/501	15.8	50/336	14.9	129/837	15.4	+ 1.0

Note that the positive differences in the last column reflect higher retention with French-identity bilinguals and NSs; conversely, negative differences reflect higher retention with English-identity bilinguals and near-natives. From these tables, we see that five speakers across both sites vary in *ne*-retention percentage (in either direction) by over 10%, the highest being 3L at 21.7% difference across interlocutor types. Eleven speakers (that is, more than half of all speakers across both sites)

differ by less than 5% across both conversations, and thirteen speakers differ by less than 8%. Thus, even though a subset of speakers produced substantial variation in both directions contributing to an overall leveling of percentages, a large majority of the speakers produced rather little intra-speaker variation in *ne*-retention across conversations.

4.4.5 *Ne*-retention results for L2 French speakers: Conclusions

As a whole, the data indicate that learners at both proficiency levels exhibit much inter-speaker variation in terms of simple frequencies of the sociolinguistic variable of *ne*-retention, much like in previous studies. SA learners produce significantly higher *ne*-retention than Near-NSs, and Near-NSs as a group produce significantly higher *ne*-retention than NSs. However, many Near-NSs overlap with NS rates.

Concerning the effect of the interlocutor, differences in interlocutors' native language backgrounds appear to strongly influence *ne*-retention rates in SA learners, with significantly lower retention in conversation with both a near-native speaker and a native speaker compared to conversation with another SA learner. Near-NSs, on the other hand, do not exhibit significant intra-speaker differences in *ne*-retention according to interlocutor L1 background.

4.5 Variationist analysis: *Ne*-retention

If negation contexts all occurred in identical linguistic and social situations, and a difference in SA learner *ne*-retention rates were found based on the interlocutor's language background, the interlocutor effect could presumably explain the differences in the SA learner *ne*-retention rates. However, an uncontrolled oral production task such as a spontaneous conversation will yield many different linguistic contexts in which negation occurs, and the data could be the result of the conversations creating imbalances in the types of contexts that favor or disfavor *ne*-

retention. A variationist analysis taking into account the influence of multiple simultaneous factors will be more instructive in determining how significant interlocutor type is, relative to other factors, in the decision to retain or omit *ne*, as opposed to random variation that yielded, for example, linguistic contexts that as a whole favored *ne*-retention in one interlocutor group over another. We can also determine the relative influence of certain characteristics related to the speaker, her interlocutor, and the context of the production task (as initially outlined in Chapter 1).

The next step is to identify the factor groups in this study that may influence variation in *ne*-retention. These factors are described in the following subsections, separated into three categories: linguistic, extralinguistic, and sociostylistic, with explanations and examples of each. For linguistic factors, the examples provided all contain a retained *ne*; however, in all cases, its deletion is also possible. All factors save interlocutor L1 have been identified in previous studies on *ne*-retention (e.g., Donaldson, 2017; Regan, 1996; Sax, 2003).

4.5.1 Linguistic factors

This section revisits the linguistic factor groups chosen for inclusion in variationist analyses of *ne*-retention, as first outlined in sections 4.2.1 and 4.3.1. Table 4-20 shows the linguistic factors that have appeared in multiple studies on *ne*-retention in L1 and L2 French. Though not an exhaustive list of variationist studies on *ne*-retention, this table is provided here as a representative sample of the factors considered to be relevant in accounting for linguistic influences on variation in *ne*-retention.

Table 4-20. Linguistic factor groups for previous variationist analysis studies on ne-retention

Factor group	Ashby (1976)	Hansen & Malderez (2004)	Coveney (2002)	Auger & Villeneuve (2008)	Regan (1996)	Sax (2003)	Meisner (2016)	Donald-son (2017)
subject type	✓	✓	✓	✓	✓		✓	✓
verb type	✓			✓	✓	✓	✓	✓
clause type	✓				✓		✓	✓
lexicalized expression	✓	✓	✓	✓	✓	✓	✓	✓
negator	✓	✓	✓	✓		✓	✓	✓
reinforcing adverb	✓	✓					✓	✓
verbal mood	✓						✓	
preceding phon. context					✓		✓	
following phon. context		✓			✓	✓	✓	
object clitic		✓	✓	✓	✓	✓	✓	
subject pronoun type		✓		✓				
subject doubling				✓				
distance: *ne*…negator				✓				

Based on these studies, the linguistic factor groups outlined below were chosen for the current study. A brief description of each factor group, along with examples of each factor, are also provided. These linguistic factors apply to both the SA learner group and the Near-NS groups.

1) Phonological environment: The phonological environment surrounding the *ne* particle is a hiatus context (that is, intervocalic position or V_V) or non-hiatus context (V_C, C_V, or C_C).

Hiatus	*Mari<u>e</u> n'<u>e</u>st pas venue.* (V 'ne' V)	'Marie didn't come.'
Non-hiatus	*Mari<u>e</u> ne <u>v</u>ient pas.* (V 'ne' C)	'Marie isn't coming.'

2) Verb type: The negated verb can be a main verb, modal, auxiliary (with *avoir* or *être*), or copula.

Main verb	*Je ne <u>parle</u> pas le chinois.*	'I don't speak Chinese.'
Modal	*Je ne <u>veux</u> pas parler le chinois.*	'I don't want to speak Chinese.'
Auxiliary	*Je n'<u>ai</u> pas <u>voyagé</u> à Paris.*	'I didn't travel to Paris.
	Je ne <u>suis</u> pas <u>allé</u> à Paris.	'I didn't go to Paris.'

Copula	*Il n'<u>est</u> pas avec moi.*	'He isn't with me.'

Modal verbs include *aller* ('go'), *pouvoir* ('be able'), *vouloir* ('want'), *devoir* ('must'), and *falloir* ('must').

3) Lexicalization: The negated element may be part of a common expression (usually learned as a "chunk" or lexicalized). For this factor, five specific lexicalized expressions were identified: *il (n')y a pas, il (n')y avait pas*, *je (ne) sais pas, c'est pas / ce n'est pas, il (ne) faut pas* ('there is/are not, there was/were not, I don't know, it's not, one shouldn't'; cf. Donaldson, 2017; Sax, 2003).

4) Post-verbal negator: There are several options for the negative adverb (or negator) following the negated verb. The standard negator *pas* ('not') is identified along with all other negators: *jamais, plus, rien, personne, aucun, ni…ni* ('never, no longer, nothing, no one, none, neither…nor'). There may be multiple such non-*pas* negators within a single clause; these tokens were coded in the 'All others' category.

pas	*Je ne parle <u>pas</u> italien.*	'I don't speak Italian.'
All others	*Je ne parle <u>jamais</u> avec mon voisin.*	'I never speak with my neighbor.'

While there has been debate over the polarity of restrictive *ne…que* ('only') (cf. Dekydtspotter, 1993; Gaatone, 1999), it is presumed that learners will treat it like negative expressions with omission of *ne* as a possibility (Dekydtspotter & Petrush, 2006). It is included in the current analysis, grouped in the category of 'All others.' Finally, *pas* can co-occur with the restrictive *ne…que* in utterances such as *ce n'est pas que lui* ('it's not just him'); due to the marked status of this structure, it is included with 'All others.'

5) Subject type: The subject of the negated verb may be a subject pronoun or a noun phrase, or the subject may not be expressed.

Subject pronoun	*<u>Je</u> ne parle pas allemand.*

'I don't speak German.'

Noun phrase	*Ma mère ne parle pas allemand.*
	'My mother doesn't speak German.'
No subject	*N'en parlons pas.*
	'Let's not talk about it.'

Subject pronouns include personal (*je/tu/il/elle/on/nous/vous/ils/elles*) and neuter (*ce/ça*) clitics, as well as the relative pronoun *qui.* Subject-doubling contexts, where a lexical NP is followed by a subject pronoun, are included with pronouns. Unexpressed subjects include imperative and infinitive structures, but not instances of phonetically deleted subjects (e.g., *faut pas* versus *il faut pas*), where *ne* would be categorically deleted.

6) Clause type: The negated clause may be a main clause or subordinate clause. Subordinate clauses include complement clauses, relative clauses, and adverbial clauses.

Main clause	*Il n'est pas poli.*
	'He is not polite.'
Subordinate clause	*Je pense qu'il n'est pas poli.*
	'I think he is not polite.'
	une personne qui n'est pas polie
	'a person who is not polite'
	Puisqu'il n'est pas poli, il me rend mal à l'aise.
	'Since he is not polite, he makes me uncomfortable.'

7) Object pronoun (clitic): The negated clause may contain one or more object pronoun clitics, appearing between *ne* position and the negated verb.

No clitic	*Je ne vois pas le livre.*	'I don't see the book.'
Clitic	*Je ne le vois pas.*	'I don't see it.'

8) Reinforcing adverb: The negated clause may contain an adverb reinforcing negation. Ashby (1976: 123) found that a reinforcing adverb favored *ne*-retention, giving two examples from his corpus: *Je ne veux absolument rien; Je ne vous comprends pas du tout* ('I want absolutely nothing; I don't understand you at all'). Based on these examples, it is clear that Ashby

considered adverbs appearing before or after the verbal negator. Meisner (2016) also considered reinforcing adverbs both before and after the negator, though she only gives examples of pre-negator adverbs with *pas.* Hanson and Malderez (2004) also coded for this factor, but they appear to have considered only adverbs that appeared between the verb and the negator (with the example *je n'aime <u>vraiment</u> pas*). The placement of the adverb can have an effect on the intended meaning; the role of an adverb as intensifier indeed occurs when it immediately follows the verb, rather than the negator. This can occur in English as well, illustrated in the following examples:

Ça ne vaut pas vraiment le coup. 'It's not really worth it.'
Ça ne vaut vraiment pas le coup. 'It's really not worth it.'

In the current study, I include all tokens of *du tout* ('at all') appearing after the negator, whether immediately (such as *c'était pas <u>du tout</u> correct* 'it was not at all correct') or within the same clause (such as *c'est pas efficace <u>du tout</u>* 'it's not efficient at all'), and reinforcing adverbs appearing between the verb and the negator, which included the adverbs *même* ('even') and *surtout* ('especially'), as well as adverbs ending in *–ment* (corresponding to the *–ly* ending in English), such as *vraiment, absolument, strictement* ('really,' 'absolutely,' 'strictly'), etc.

4.5.2 Commentary on additional linguistic factors

Compared with the studies in Table 4-20, several linguistic factor groups in the current study were modified or excluded from consideration and merit discussion in this section. First, an initial analysis, limited to the SA learner group, included phonological environment as two separate factor groups (preceding and following phonological segment, with different factors for consonants and vowels). Donaldson (2017) considers the factor of phonological environment itself to be problematic, not only due to inconsistent results in previous studies, but also due to the fact

that, citing Coveney (2002: 78), the presence or absence of *ne* "radically" alters the phonological context because of assimilation and elision processes. Elsewhere, Coveney (1998: 167) summarizes this problem as follows: "It seems that…the [*ne*] variable is affecting the phonological environment, rather than vice versa." For SA learners in the current study, the factor of following phonological segment was revealed to be marginally significant in the preliminary analysis; SA learners slightly favored *ne*-retention when the segment was consonantal. Due to its marginal significance, and the problems mentioned in previous studies, I modified this factor in the current study to analyze a more restrictive phonological context: rather than all contexts before and after *ne* (or its expected contexts when absent), I limited the phonological context only to those situations concerning a hiatus of intervocalic position (consecutive oral or nasal vowel segments[51]). The insertion of *ne* is thus a single phonological segment, [n], and its presence or absence does not change (that is, cause elision of) the preceding or following vowel. For example, in *ce n'est pas* ('it is not'), *ne*-deletion would not result in hiatus, due to elision, *c'est pas* versus **ce est pas*. In general, hiatus contexts tend to be avoided in French, so the opportunity to avoid hiatus by insertion of *ne* may favor its retention (cf. Ashby, 1976: 129; Pohl, 1968: 1354). I also did not consider tokens containing the string *tu n'es pas* ('you are not') or *tu n'as pas* ('you have not') as potential hiatus contexts, since it is impossible to determine whether *ne*-deletion in the same context would result in a hiatus in informal speech (*t'es pas* versus *tu es pas*; *t'as pas* versus *tu as pas*). In the corpus, only one token of deletion resulting in hiatus was produced (the native speaker SA interlocutor (SoF): *Tu as pas fait Paris?* 'You didn't do Paris?'); in general, native and

[51] I follow Meisner (2016) in combining preceding and following nasal vowels into this category of hiatus (thus including Ṽ_V and V_Ṽ along with consecutive oral vowels V_V). Meisner's corpus contains three tokens of nasal vowels in this phonological environment (of which two retained *ne*) along with 12 tokens of V_V. My corpus contains 10 tokens of nasal vowels, with one *ne* omitted (by bilingual speaker Ch in the V_Ṽ utterance *toi qui en voulais pas* 'you who didn't want any').

near-native speakers elided the /y/ vowel in *tu* when followed by vowel-onset verbs, even in negation contexts: *t'es pas* /tepa/. This supports the general tendency to avoid hiatus, even though it can occur in more careful styles (e.g., *tu es étudiant ?* 'you're a student?').

Moving to other linguistic factors, Ashby (1976) found evidence that the presence of a second negator increased *ne*-retention. This was not borne out in the current study, with *ne* being retained in two of 22 occurrences of multiple negators (one *ne* with *pas que* and one with *Personne…plus*, both uttered by Near-NSs). All multiple negators produced are listed as follows: *Personne…plus, Personne…jamais, pas que* (11 occurrences), *plus personne, plus rien* (3 occurrences), *plus jamais* (2 occurrences), *plus aucune, jamais personne*, and one triple negator, *plus jamais rien*. Note that the SA learners also produced two ungrammatical uses of multiple negators, both overapplying *pas* with *personne*: 1) *je n'ai pas connu personne* ('I didn't know anyone'; cf. grammatical *je n'ai connu personne*); 2) *personne ne me dit pas* ('no one tells me'; cf. grammatical *personne ne me dit*).

Coveney (2002) and Hansen and Malderez (2004) also cite differences in nonfinite versus finite verbs, where non-finite verbs would favor *ne*-retention.

Finite	*Je ne travaille pas à Paris.*	'I'm not working in Paris.'
Nonfinite	*...pour ne pas travailler à Paris.*	'…in order not to work in Paris.'

Note that the word order of negation changes in non-finite constructions. Other studies have excluded this factor due to this syntactic inversion of the negator and the verb. In the current study, I included nonfinite verbal negation but did not code for this factor separately in the variationist analysis, as such structures can be accounted for via the factor group of subject type, where negation with no subject includes nonfinite verbs along with imperatives.

4.5.3 Extralinguistic factors: SA learners

Due to the rather homogeneous nature of the SA learner group, extralinguistic factors that often appear in other variationist analyses can be excluded for this group, including sociobiographical factors such as age, occupation/education level, study-abroad status, and country of origin. Since all oral tasks were conducted in the same conversational setting, formality was also excluded as a factor.

1) Gender (male, female). Uncommonly for a group of L2 French learners, males outnumbered females. While the native and near-native interlocutors were both the same gender, this factor was not controlled in the learner group.
2) Interlocutor L1 (English, French). For this factor, the near-native speaker of French is included with the learner group.
3) Interlocutor type (NS, near-native, SA learner). Since the analysis of simple *ne*-retention frequencies showed differences across types of interlocutors, each type will be considered as a separate factor, thus separating the near-native interlocutor from other SA learners.
4) Conversation portion: Speakers may become more comfortable later in the conversation compared to the beginning, due to various factors such as the recording environment and the speaker's relationship with the interlocutor, with possible effects on stylistically influenced variation. Conversations are divided into three portions (cf. Auger & Villeneuve, 2008): first 5 minutes; 5-10 minutes; after 10 minutes.
5) Speaker: Each speaker may be a significant source of variation (Regan, 1996). As in previous studies (Donaldson, 2017), each speaker was selected as a random intercept in the Rbrul analysis.

4.5.4 Extralinguistic factors: Near-NSs

Due to the modified nature of the tasks and the different overall profile of the Near-NSs, the extralinguistic factors for this learner group do not entirely overlap with those in the SA learner group. The following factors included for analysis for Near-NSs are explained as follows:

1) Interlocutor L1: For Near-NSs in Lille, the interlocutor is a native or non-native speaker of French. For Near-NSs in Pau, the interlocutor has adopted an L1 French or L1 English identity.
2) Age group: As many L1 studies have done, we can divide, post-hoc, the Near-NSs into two age groups: nine younger speakers (age range = 26-38) and ten older speakers (age range = 44-75). Though the age cut-off is certainly an arbitrary choice, it divides the Near-NSs into two nearly even groups, and the number of years (6) between the oldest "young" speaker (38) and the youngest "old" speaker (44) is larger than the gap between any of the ages in the younger group and all but one of the gaps in age in the older group (where there is a gap of nine years between two speakers aged 53 and one aged 62).
3) Length of residence (LOR): Similar to age, a separation of the Near-NSs can be made between those who have lived in France for at least 10 years compared with those who have lived in France for fewer than 10 years. If speakers with longer LOR as a whole have significantly lower *ne*-retention than those with shorter LOR, it suggests that *ne*-retention rates may not yet be stable in speakers living in France fewer than 10 years. Note, however, that for Near-NSs in Lille, the three speakers who have lived in France for at least 10 years are also the three speakers in the older age group for this site. Since the Length of residence factor group overlaps completely with the Age factor group for this site, the former factor group will not be included for Near-NSs in Lille.

4) Gender (male, female). This factor was more balanced for Near-NSs in Lille (four males, five females) than in Pau (one male, nine females).
5) Conversation portion
6) Speaker. This factor is coded as a random intercept to account for inherent random variation among speakers (cf. Regan, 1996).

4.5.5 Extralinguistic factors: Interlocutors

Due to the nature of the interlocutor role, with only two or three interlocutors per site, fewer of the extralinguistic factors tested for L2 speakers apply here. Only the Conversation portion is coded for variation in all speakers. NSs in Pau are coded according to the identity (English or French) they adopted for each conversation. When NSs are combined at both sites, it is possible to add gender as a factor, with three males and two females. Finally, each speaker is coded as a random intercept.

4.5.6 Sociostylistic factors

Sociostylistic factors are an example of "micro-style variation" (Armstrong 2002: 171). Contra sociobiographical factors, these factors are not inherent characteristics of the speaker, and contra linguistic factors, they are not linked at the phonological or syntactic level. Rather, they concern differences either at the level of the utterance or at the level of the context of communication, such as whether the speaker is quoting another person (or herself)—where the speaker may enact a temporary style differing in formality from the current conversation (Coveney, 2002; Poplack & St-Amand, 2007)—or whether the choice of topic is "serious," which can cause speakers to more consciously self-monitor their speech. Similarly, the formality of the context of communication can be included as a factor, whether this is manipulated as part of the

speaking task (e.g., Sax, 2003) or is determined by features such as pronouns of address (e.g., Coveney, 1996). All of these factors can yield contexts favorable to the production of the formal sociolinguistic variant, viz. retention of *ne.*

All speakers were considered for sociostylistic factors, following Donaldson (2017) and others. The SA learners were not coded for factors 1 and 3 below, due to the small number of tokens for quoted speech (five tokens, including three *ne*-retention) and for emphasis (four tokens, including three *ne*-retention).

1) Quoted speech: Does the negated clause contain quoted speech, and if so, is the speaker quoting from a more formal style? The identification of a more formal style is generally easy to discern, given the conversational context (e.g., quoted speech containing the formal singular address pronoun *vous*, or contextual clues such as quoting speech to one's supervisor). Since the current study's Near-NS participants included several teachers, there were a number of quoted speech utterances of students speaking to teachers and vice-versa, as well as hypothetical or general attitudes of students or teachers in which the speaker adopted a subject pronoun (e.g., *je*, *vous*) indicating that the speaker was quoting real or supposed speech. I considered student speech directed to teachers as quoting a formal style (generally supported by the use of the formal singular *vous*); given the context of a classroom setting, I also considered teacher speech directed to students as quoting a formal style, even though the latter speech generally included the informal subject pronoun *tu* (or subject-less imperatives with the *tu* inflection). In coding negation tokens for this factor group, I used three categories: *not quoted*, *quoted speech – formal*, and *quoted speech – not formal.*
2) Serious discourse topic: Serious topics that have been identified in previous literature include discussions about language/metalanguage, religion/sermons, education, meetings, moralizing,

one's profession, discipline of children, and the legal system (Coveney, 2002; Fonseca-Greber, 2007; Pohl, 1975; Poplack & St-Amand, 2007; Rehner & Mougeon, 1999; Sankoff & Vincent, 1977; van Compernolle, 2009). In sociolinguistic interviews, such topics can be selected by the interviewer beforehand (e.g., Rehner & Mougeon, 1999); otherwise, in free conversation tasks, the researcher can retroactively make a determination as to whether the speech sample contains topics that are serious or not (e.g., Donaldson, 2017). Due to the design of the current study, I have adopted the latter approach. I have also simplified topic formality compared to previous studies (e.g., Rehner & Mougeon, 1999) and adopted a binary coding scheme as either *serious* or *neutral/informal* topics. Concerning SA learners, the topic was somewhat controlled in that the dyads were provided with suggested topic prompts for conversation, though they were free to deviate; SA learners tended to use the provided prompts more with other learners than with the native/near-native interlocutors. The topic of "education" was also problematic, as much of the conversation concerning SA learners' daily routines and background involved discussion of their classes and their studies at their home universities. For SA learners, this topic was not included in the "serious" category. Most of the remaining "serious" topics included discussion on metalanguage.

3) Emphasis: Is the negation emphasized in a particular way? According to Fonseca-Greber (2007), emphatic negation can include the following contexts: lexical emphasis, repeated speech, slower speech, pitch prominence, and contrast, all of which may favor the retention of *ne.* Donaldson (2017) considered all such emphatic contexts his analysis of *ne*-retention, adopting a binary coding scheme as "emphatic speech" or not; however, given the non-native status of his speakers, "emphatic speech" did not include pauses or slower speech due to word search hesitations. Despite the subjective nature of determining a threshold for considering an

utterance as emphatic, I followed Fonseca-Greber's analysis as closely as possible while excluding word search hesitations as did Donaldson, and I adopted a binary coding scheme with the labels *emphasis* and *no emphasis.*

4) *Tu* versus *vous*: Despite the fact that these conversations were designed to be rather informal (no interviewer versus interviewee roles, location of recording designed to be casual, and my general instruction to speak on any topic that came to mind), the methodology required a certain level of artificiality, such as the presence of recording equipment (though minimized as much as possible), and the fact that none of the participants, except for the SA learner-SA learner dyads, had known each other before the start of the speaking task. Each speaker was therefore free to adopt a more formal or more informal tone with her interlocutor; I did not instruct the participants on which pronouns of address to use.[52] Differences in social distance created by the adoption of either pronoun in a given conversation could influence the use of stylistically-conditioned variables such as *ne*-retention, so its potential effect should be considered. As noted in Chapter 3, a review of the recordings revealed that both *tu* and *vous* were used in the SA learner and Near-NS proficiency groups, by learners and native speakers alike (though the native interlocutor and near-native interlocutor for the SA learners used *tu* with all eight learners). Note also that in two of the SA learner conversations, the learner used no personal address pronouns with the interlocutor, so negation tokens from these conversations were placed in a third, "unknown formality" category.

[52] Coveney (1996) likewise did not instruct his conversation partners on the use of *tu/vous*; 24 of his conversations were conducted using *tu.*

4.5.7 Results: SA learners

Returning to the current study, I begin with the variationist analysis for the SA learner group. All participants were coded for the factor groups as described earlier in this section, with the dependent variable *ne*-retention, and I used Rbrul (Johnson, 2009) to carry out a variationist analysis. Recall that two participants in the SA learner group (6S and 7S) produced near categorical *ne*-retention, with nearly identical frequencies and number of tokens (6S: 51/52, 98.1%; 7S: 50/51, 98%), and recall that these speakers' lone *ne*-deletion tokens were considered marginal, due to a pause or immediate repetition (refer to section 4.4.2 for details). Based on these issues, I excluded these two speakers from the variationist analysis. This limits the generalizability of the analysis somewhat (with six SA learners remaining), but the number of speakers and tokens is nevertheless sufficient for a variationist treatment of the data.[53]

Table 4-21 provides details on the factor groups selected by Rbrul as significant in explaining the variable usage of *ne*. For each factor group, all coded factors are listed. Log-odds coefficients above zero and factor weights above .5 indicate that the factor favors *ne*-retention, while negative log-odds and factor weights below .5 indicate a disfavoring effect. Earlier variationist analysis models such as Goldvarb X rely on factor weights to represent the favoring or disfavoring of a factor on the variable; log-odds are used in Rbrul to better represent effect sizes. The raw percentages of *ne*-retention for each factor are also provided; in the last column, *N* indicates the number of *ne*-retention tokens out of all negation tokens. Furthermore, note that the factor groups are listed in decreasing order of significance. Thus, Lexicalization is the most significant factor group accounting for variation in SA learners, and Interlocutor type is the least significant of the remaining significant factor groups. The appearance of Interlocutor type as

[53] The *ne*-retention rates (223/391, 57.0%) for the remaining speakers showing variation is quite similar to Sax (2003), 60.8%, providing a useful cross-sample comparison.

significant is nevertheless substantial, given Preston's (1991) observation of a primacy of linguistic factors over non-linguistic factors (cf. section 2.10).

Below each table, I provide the input probability, which is an average of the predicted values for each cell, providing a baseline for the model. I also provide the log likelihood, which measures the goodness of fit of each analysis; if this value is closer to zero, it represents a better model (with a range from 0 to -1000). Finally, I list the non-significant factor groups for each table; for more statistical details on these non-significant factors for all Rbrul tables, see the corresponding tables in Appendix G.

Table 4-21. Significant factor groups for SA learners (ne-retention)

Factor Group	**Factor**	**Log odds**	**Factor weight**	**% *ne*-retention**	**N**
Lexicalization	not lexicalized	0.549	.634	73.6	142/193
	lexicalized	-0.549	.366	40.9	81/198
Verb type	auxiliary	0.974	.726	85.7	18/21
	main	0.862	.703	72.8	171/235
	modal	-0.169	.458	57.9	11/19
	copula	-1.666	.159	19.8	23/116
Hiatus	hiatus	1.178	.765	88.9	8/9
	no hiatus	-1.178	.235	56.3	215/382
Interlocutor type	SA learner	0.524	.628	74.5	82/110
	NS	-0.229	.443	49.3	70/142
	Near-NS	-0.295	.427	51.1	71/139
TOTAL				57.0%	223/391

Input probability = 0.808; Log likelihood = -192.496
Non-significant factor groups (cf. Table G-1): Object clitic, Clause type, Reinforcing adverb, Subject type, Negator, Topic, *Tu/vous*, Interlocutor L1, Gender, Conversation portion

As noted, Lexicalization is the most significant factor group accounting for *ne* usage in this learner population, with lexicalized expressions disfavoring *ne*-retention (indicated by negative log-odds and a factor weight below .5), repeating Regan's (1996) findings for linguistic factors in L2 speakers. Verb type was also significant in the current study, with auxiliary and main verbs favoring *ne*-retention; this factor was only marginally significant in Sax (2003) and not significant

in Regan (1996). Unsurprisingly, copulas disfavored *ne*-retention, as Sax also found. Despite low token counts, learners at this proficiency level also demonstrate sensitivity to phonological hiatus; recall that in Regan (1996) and Sax (2003), such sensitivity was confirmed only for phonological segments following *ne*. Concerning extralinguistic factors, it is crucial that Interlocutor type appears as a significant factor, with *ne*-retention favored in conversations with fellow SA learners and disfavored when conversing with native and near-native speakers; indeed, it is the only significant extralinguistic factor.

As for non-significant factor groups (cf. Table G-1 in Appendix G), despite small token counts for some factors, we see some impact of these factors aligning with observations in previous studies, even if they are not significant in the variationist analysis, such as full NPs favoring *ne*-retention over subject clitics (cf. Ashby, 1976; Regan, 1996). Other factors, however, trend in the opposite direction: higher *ne*-retention later in the conversation (contra Ashby, 1976), and higher *ne*-retention with object clitics (contra Regan, 1996).

4.5.7.1 SA learners: Commentary

Despite the small sample size, learners in this study clearly exhibit specific patterns concerning *ne*-retention. As expected, *ne*-retention for all SA learners is much higher than that produced by the native and near-native interlocutor, as well as that measured by native speakers in previous studies. Crucially, these learners show significant differences across interlocutor type, with native/near-native conversations tending to disfavor *ne*-retention in comparison with conversations with another SA learner. These results appear to demonstrate learner convergence, both with speakers at a higher proficiency level and with a speaker at or near the same proficiency level. In his study, Dewaele (2004a: 445) explains this convergence of learners in interaction with native speakers as follows: "[The learners] were temporarily cast in the role of L2 learners rather

than legitimate L2 users. They might have interpreted this particular social interaction as another learning experience; hence their desire to sound as much as possible like the NSs they were talking to, and converging maximally." It can be postulated that SA learners in the current study, while capable of more nativelike speech with higher-proficiency speakers, also converge with other learners, whether this is due to more conscious monitoring of their speech,[54] or simply due to intense social pressure to conform to the speech patterns of their peers. Regardless, the increased social convergence of the learner-learner dyad (specifically, study-abroad student with study-abroad student) did not lead to style-shifting in an informal direction; rather, it had the opposite effect.

It is noteworthy that what heavily influences *ne*-retention in SA learners are two linguistic factors focusing on verbal structures (an unsurprising finding, given the status of *ne* as a clitic with a syntactically strong attachment to the finite verb). Returning to the status of the NS/near-native interlocutors, in the current study, it appears that SA learners treated the near-native much like a native speaker. However, in the debriefing session, only one student revealed being unaware that the near-native was in fact not a native speaker. At least for these study-abroad learners, the actual interlocutor language background is only part of the story: distance in (perceived) proficiency also appears to influence convergence toward native speaker (and nativelike speaker) norms. This may explain why Interlocutor type (as native, near-native, or SA learner) was selected as a significant factor group, rather than the Interlocutor L1 factor with its binary status as native or non-native (grouping the near-native and SA learners together).

[54] In the debriefing session, all learners remarked that they were aware of *ne* as a marker of formality; while some indicated that they did pay attention to how they used *ne* in their own speech in general, none reported having actively monitored its use during these conversations.

4.5.7.2 Results: Interlocutors with SA learners

As for the native interlocutor and near-native interlocutor, a variationist analysis of their *ne*-retention in conversation with SA learners revealed only three significant factor groups (Table 4-22).

Table 4-22. Significant factor groups for SA interlocutors (ne-retention)

Factor Group	Factor	Log odds	Factor weight	% *ne*-retention	N
Emphasis	emphasized	1.725	.849	75.0	3/4
	not emphasized	-1.725	.151	6.2	19/306
Lexicalization	not lexicalized	0.742	.677	11.3	19/168
	lexicalized	-0.742	.323	2.1	3/142
Subject type	NP	1.522	.821	66.7	2/3
	pronoun	-1.522	.179	6.5	20/307
TOTAL				7.1%	22/310

Input probability = 0.538; Log likelihood = -66.264
Non-significant factor groups (cf. Table G-2): Hiatus, Object clitic, Verb type, Clause type, Reinforcing adverb, Negator, Topic, Quoted speech, Conversation portion

The lower number of tokens and the low overall *ne*-retention rate may account for the selection of few factor groups for these two speakers. Lexicalization is a clear factor influencing *ne*-retention, with lexicalized expressions strongly disfavoring retention (of the three tokens with retention, two were the expression *je n'sais pas* 'I don't know') by near-native AmE and one was the expression *il n'y avait pas* ('there wasn't') by native SoF). Otherwise, when looking at the results from the non-significant factor groups (cf. Table G-2), these factor groups generally concord with previous findings regarding *ne*-retention: favoring retention in hiatus contexts, subordinate clauses, negators other than *pas*, and serious topics (though there was categorical deletion with reinforcing adverbs). However, these factor groups were not selected as significant in the variationist analysis, due either to small token counts, smaller differences across individual factors, or a combination of both.

4.5.8 Results: Near-NS groups and interlocutors

I now turn to the variationist analysis obtained for the Near-NSs and their interlocutors. Using the factors for Near-NSs outlined in previous sections, tokens for Near-NSs and their interlocutors were coded in Rbrul, and a variationist analysis was conducted for certain groupings of speakers: Near-NSs in Lille, Near-NSs in Pau, Near-NSs grouped all together, NSs at each site, NSs grouped all together, and near-native interlocutors in Lille. The following tables provide details on the significant factor groups for each grouping of speakers, along with a list of non-significant factors; see Appendix G for more detailed tables on non-significant factors. I begin with the results for Near-NSs in Lille (Table 4-23) and Near-NSs in Pau (Table 4-24).

Table 4-23. Significant factor groups for Near-NSs in Lille (ne-retention)

Factor Group	Factor	Log odds	Factor weight	% *ne*-retention	N
Hiatus	hiatus	1.657	.840	81.8	18/22
	no hiatus	-1.657	.160	13.6	111/815
Emphasis	emphasized	1.155	.760	53.8	21/39
	not emphasized	-1.155	.240	13.5	108/798
Verb type	auxiliary	0.914	.714	24.4	22/90
	main	0.500	.622	19.6	80/409
	copula	-0.322	.420	8.9	24/271
	modal	-1.093	.251	4.5	3/67
Topic	serious	0.406	.600	20.8	51/245
	neutral/informal	-0.406	.400	13.2	78/592
Subject type	none	0.801	.690	60.0	3/5
	NP	0.542	.632	43.8	7/16
	pronoun	-1.343	.207	14.6	119/816
Reinforcing adverb	no adverb	0.638	.654	15.7	123/784
	adverb	-0.638	.346	11.3	6/53
TOTAL				15.4%	129/837

Input probability = 0.705; Log likelihood = -280.149
Non-significant factor groups (cf. Table G-3): Object clitic, Clause type, Lexicalization, Negator, Quoted speech, Interlocutor L1, Age, Gender, Conversation portion

Table 4-24. Significant factor groups for Near-NSs in Pau (ne-retention)

Factor Group	Factor	Log odds	Factor weight	% *ne*-retention	N
Subject type	NP	1.035	.738	78.7	48/61
	none	0.938	.719	68.2	15/22
	pronoun	-1.974	.122	16.6	189/1144
Hiatus	hiatus	1.431	.807	78.9	45/57
	no hiatus	-1.431	.193	17.7	207/1167
Emphasis	emphasis	0.869	.704	59.3	35/59
	no emphasis	-0.869	.296	18.6	217/1165
Lexicalization	not lexicalized	0.561	.637	25.9	228/879
	lexicalized	-0.561	.363	7.0	24/345
Verb type	auxiliary	0.590	.643	31.8	34/107
	main	0.185	.546	22.1	145/656
	copula	-0.240	.440	16.6	54/326
	modal	-0.535	.369	14.1	19/135
TOTAL				20.6%	252/1224

Input probability = 0.829; Log likelihood = -422.987
Non-significant factor groups (cf. Table G-4): Object clitic, Clause type, Reinforcing adverb, Negator, *Tu/vous*, Topic, Quoted speech, Interlocutor L1, Age, Length of residence, Gender, Conversation portion

Common significant factor groups for Near-NSs at both sites include Hiatus, Verb type, Emphasis, and Subject type. Though the tokens for hiatus in Lille are few, the greater number in Pau and the nearly identical *ne*-retention rates at both sites strengthen the overall effect observed for this factor, where a hiatus context favors *ne*-retention. Results for Subject type pattern as in previous studies, with NPs favoring retention and pronouns disfavoring retention, though the Near-NSs in Lille have a comparatively low retention rate with NPs (albeit from only 16 tokens). Lexicalized expressions are not significant in Lille, as Donaldson (2017) found for his near-natives (though Donaldson found individual variation in this group concerning sensitivity to this factor, which was significant at the group level for Near-NSs in Pau). As for sociostylistic factors, of the three significant factors influencing *ne*-retention in Donaldson's near-native speakers, Emphasis is the only one that was significant at both Lille and Pau, with emphatic contexts favoring *ne*, though Topic and Quoted speech were both marginally insignificant for Near-NSs in Pau.

When all Near-NSs are combined in an additional Rbrul run, all factor groups that were significant at each site remain significant overall, save for Reinforcing adverb, as Table 4-25 shows.

Table 4-25. Significant factor groups for Near-NSs in Lille and Pau combined (ne-retention)

Factor Group	Factor	Log odds	Factor weight	% *ne*-retention	N
Subject type	none	1.448	.810	66.7	18/27
	NP	1.342	.793	71.4	55/77
	qui	-1.322	.210	49.2	30/61
	other pronoun	-1.468	.187	14.7	278/1896
Hiatus	hiatus	1.399	.802	79.7	63/79
	no hiatus	-1.399	.198	16.0	318/1982
Emphasis	emphasized	0.959	.723	57.1	56/98
	not emphasized	-0.959	.277	16.6	318/1963
Verb type	auxiliary	0.673	.662	28.4	56/197
	main	0.284	.571	21.1	225/1065
	copula	-0.297	.426	13.1	78/597
	modal	-0.659	.341	10.9	22/202
Lexicalization	not lexicalized	0.342	.585	22.7	329/1451
	lexicalized	-0.342	.415	8.5	52/610
Topic	serious	0.237	.559	23.6	144/611
	neutral/informal	-0.237	.441	16.3	237/1450
Object clitic	no clitic	0.378	.593	18.5	364/1963
	clitic	-0.378	.407	17.3	17/98
TOTAL				18.5%	381/2061

Input probability = 0.734; Log likelihood = -709.963
Non-significant factor groups (cf. Table G-5): Clause type, Reinforcing adverb, Negator, Quoted speech, *Tu/vous*, Interlocutor L1, Age, Length of residence, Gender, Conversation portion, Site

Lexicalization, which was significant for Pau but not Lille, remains significant for the combined run. Also note that Object clitic was selected as significant (though only marginally so, p = .033), with a pre-verbal clitic disfavoring retention as previous studies have found for L1 speakers (cf. Posner, 1985; Larrivée, 2014; Meisner, 2016). In addition to Emphasis, Topic remains significant here, with serious topics slightly favoring *ne* and neutral or informal topics slightly disfavoring it, while Quoted speech was marginally insignificant in the combined run.

These Near-NSs thus show similar sensitivity to the three sociostylistic factors as did Donaldson's group of near-natives. However, these Near-NSs show sensitivity to more linguistic factors than Donaldson's group, who showed sensitivity only to Subject type and Relative clauses.

For the NS interlocutors, Rbrul runs were conducted at each site and then combined. Table 4-26 (Lille) and Table 4-27 (Pau) provide the details on significant factor groups for these NSs.

Table 4-26. Significant factor groups for NSs in Lille (ne-retention)

Factor Group	Factor	Log odds	Factor weight	% *ne*-retention	N
Hiatus	hiatus	2.089	.890	70.6	12/17
	no hiatus	-2.089	.110	8.6	53/617
Subject type	NP	1.810	.859	54.5	6/11
	none	0.123	.531	33.3	5/15
	pronoun	-1.933	.126	8.9	54/608
Emphasis	emphasized	1.156	.761	40.6	13/32
	not emphasized	-1.156	.239	8.6	52/602
Interlocutor ID	4L	1.815	.860	28.8	23/80
	7L	0.543	.633	12.0	6/50
	5L	0.431	.606	8.7	8/92
	3L	-0.081	.480	7.5	6/80
	6L	-0.122	.469	5.3	3/57
	1L	-0.188	.453	8.8	7/80
	2L	-0.261	.435	5.4	5/93
	9L	-0.263	.435	6.7	3/45
	8L	-1.875	.133	7.0	4/57
Verb type	auxiliary	1.008	.733	17.2	10/58
	main	0.564	.637	12.1	46/379
	modal	-0.067	.483	5.3	3/57
	copula	-1.505	.182	4.3	6/140
TOTAL				10.3%	65/634

Input probability = .845; Log likelihood = -145.976
Non-significant factor groups (cf. Table G-6): Object clitic, Clause type, Reinforcing adverb, Lexicalization, Negator, Topic, Quoted speech, Conversation portion

Table 4-27. Significant factor groups for NSs in Pau (ne-retention)

Factor Group	Factor	Log odds	Factor weight	% *ne*-retention	N
Hiatus	hiatus	1.888	.869	50.0	14/28
	no hiatus	-1.888	.131	5.5	68/1261
Subject type	none	1.181	.765	33.3	2/6
	NP	0.965	.724	37.3	13/35
	pronoun	-2.145	.105	5.4	68/1248
Interlocutor ID[55]	1P	1.986	.879	30.1	31/103
	2P	0.732	.675	9.1	8/88
	8P	0.720	.673	3.4	5/149
	7P	0.648	.657	4.9	6/122
	5P	0.627	.652	3.9	5/129
	10P	0.346	.586	1.9	3/159
	9P	-0.179	.455	1.5	2/133
	4P	-1.399	.198	5.3	5/94
	6P	-1.565	.173	1.2	2/161
	3P	-1.918	.128	10.6	16/151
Topic	serious	0.753	.680	12.5	45/361
	neutral/informal	-0.753	.320	4.1	38/928
Emphasis	emphasized	1.077	.746	27.6	16/58
	not emphasized	-1.077	.254	5.4	67/1231
Clause type	subordinate	0.674	.662	17.4	31/178
	main	-0.674	.338	4.7	52/1111
Tu/vous[56]	*vous*	1.643	.838	7.6	79/1044
	tu	-1.643	.162	1.6	4/245
Negator	other	0.450	.611	9.7	19/196
	pas	-0.450	.389	5.9	64/1093
Object clitic	clitic	0.535	.631	9.2	8/87
	no clitic	-0.535	.369	6.2	75/1202
TOTAL				6.4%	83/1289

Input probability = 0.645; Log likelihood = -164.064
Non-significant factor groups (cf. Table G-7): Reinforcing adverb, Verb type, Quoted speech, Conversation portion, Speaker guise

For these speakers, there are several similarities to the Near-NS groups. Hiatus and Subject type remain the most significant factors across all Near-NS and NS groups; NPs, non-subjects, and hiatus again favor *ne*-retention. Verb type appears in Lille but not in Pau, with modals and copula

[55] The VIF for this factor group was above 7.5, suggesting that these values are correlated with another predictor.
[56] Rbrul reported a VIF of at least 2.5 for the factor weights in the *tu/vous* group.

être disfavoring retention as found in previous studies. NSs in Pau, however, show sensitivity to several more linguistic factors. The type of negator first appears here, with non-*pas* negators favoring retention at a marginally significant rate ($p = .034$). Object clitic is also significant, though in the opposite direction of the Near-NSs combined, with the presence of a clitic favoring retention (at a marginally significant rate, $p = .041$). Finally, as found in previous studies (Ashby, 1976; Donaldson, 2017), Clause type is significant for Pau NSs, with subordinate clauses favoring retention here.

Concerning extralinguistic factors, though Donaldson's NSs showed sensitivity to three sociostylistic factors just as his near-natives did (Quoted speech, Emphasis, and Topic), Topic was significant only in Pau, and Quoted speech appears in neither group. The *Tu/vous* factor group also appears as significant for NSs in Pau, with *vous* favoring *ne* and *tu* disfavoring it, suggesting that the pronoun of address (or the level of formality associated with this choice of pronoun) had an influence on these speakers' *ne*-retention. Finally, Interlocutor ID (referring to the Near-NSs as interlocutors) appears in both groups, which may not be surprising given the potential for variation that can occur when sampling across nine or ten different conversations; an increase in retention in a single conversation (such as bilingual Th's *ne*-retention with Near-NS 1P) can impact this factor goup more strongly than for other factor groups. Its high collinearity with other factor groups in Pau (such as *tu/vous*) also reduces the significance of this factor alone, which may also explain the unequal relationship between *ne*-retention percentages and log-odds/factor weights for certain entries in this group. Nevertheless, one cannot rule out the possibility that individual Near-NSs may explain some of the variation in NS behavior.

With this variation in mind, I conducted an additional run combining NSs at both sites, with both Speaker ID and Interlocutor ID (now with 19 total factors) as random intercepts. In the

combined run, a slightly different arrangement of significant factor groups is retained, as Table 4-28 shows.

Table 4-28. Significant factor groups for NSs in Lille and Pau combined (ne-retention)

Factor Group	Factor	Log odds	Factor weight	% *ne*-retention	N
Subject type	NP	1.287	.784	41.3	19/46
	none	0.606	.647	33.3	7/21
	pronoun	-1.893	.131	6.6	122/1856
Hiatus	hiatus	1.628	.836	57.8	26/45
	no hiatus	-1.628	.164	6.5	122/1878
Emphasis	emphasized	1.126	.755	32.2	29/90
	not emphasized	-1.126	.245	6.5	119/1833
Clause type	subordinate	0.501	.623	18.4	48/261
	main	-0.501	.377	6.0	100/1662
Topic	serious	0.377	.593	12.8	69/541
	neutral/informal	-0.377	.407	5.7	79/1382
Tu/vous	*vous*	0.877	.706	7.6	79/1044
	tu	-0.877	.294	7.8	69/879
Verb type	auxiliary	0.721	.673	12.7	23/181
	main	0.098	.524	8.1	90/1112
	modal	-0.335	.417	5.4	9/167
	copula	-0.485	.381	5.6	26/463
Quoted speech	formal	1.447	.810	29.4	5/17
	not quoted	-0.433	.394	7.7	142/1839
	not formal	-1.015	.266	1.5	1/67
TOTAL				7.7%	148/1923

Input probability = 0.887; Log likelihood = -335.019
Non-significant factor groups (cf. Table G-8): Object clitic, Reinforcing adverb, Lexicalization, Negator, Gender, Conversation portion

As with nearly all other speaker groupings, Subject type and Hiatus remain the most significant factor groups; NPs, non-subjects, and hiatus contexts strongly favor retention. The only other significant linguistic factor, Verb type, is retained here as it was in Lille, with auxiliary and main verbs slightly favoring retention. As for extralinguistic factors, all three sociostylistic factors significant in Donaldson's groups are significant when combining NSs at both sites, and Quoted speech appears here with the expected distribution: quoted formal speech favors retention, while

quoted informal/neutral speech disfavors retention. One surprising extralinguistic factor group here is *Tu/vous*, which becomes significant in the combined run, even though the difference in retention rates when combining both sites is very minimal (with a slight trend in the opposite direction of previously observed behavior; *ne*-retention with *tu* is slightly higher).

Finally, the smaller set of *ne*-retention contexts analyzed in Rbrul for the two near-native interlocutors (SaE and JeE) in Lille is reported in Table 4-29.

Table 4-29. Significant factors for near-native interlocutors in Lille (ne-retention)

Factor Group	**Factor**	**Log odds**	**Factor weight**	**% *ne*-retention**	**N**
Subject type	NP	2.445	.920	85.7	6/7
	none	-0.603	.354	33.3	2/6
	pronoun	-1.842	.137	10.7	33/308
Hiatus	hiatus	1.584	.830	83.3	5/6
	no hiatus	-1.584	.170	11.4	36/315
Emphasis	emphasis	0.931	.717	50.0	8/16
	no emphasis	-0.931	.283	10.8	33/305
Object clitic	no clitic	1.005	.732	13.9	40/288
	clitic	-1.005	.268	3.0	1/33
TOTAL				12.8%	41/321

Input probability = 0.745; Log likelihood = -98.258
Non-significant factor groups (cf. Table G-9): Relative clause, Reinforcing adverb, Verb type, Lexicalization, Negator, Topic, Quoted speech, Interlocutor ID, Conversation portion

As with the SA interlocutors, small token counts likely account for fewer significant factor groups compared with the Near-NS and NS groups, though resulting in a log likelihood closer to zero and thus a better fitting model. Subject type and Hiatus consistently appear as the most significant factors, with non-pronouns and hiatus contexts (in low token counts) strongly favoring *ne*-retention. The sociostylistic factor Emphasis was significant as found with most Near-NS and NS groupings, with emphatic contexts strongly favoring *ne*; in contrast, Topic, and Quoted speech were not selected as significant. Finally, object clitics disfavor retention, as found in the combined Near-NS grouping, though the difference in retention was much greater for these interlocutors.

4.5.9 Near-NS groups and interlocutors: Commentary

This section provides further discussion of the significant factor groups as identified in the preceding variationist analyses for the Near-NSs and their interlocutors. Overall, more factors were identified as significant compared with the SA learner group, with Subject type, Verb type, and Hiatus contexts, as well as several sociostylistic factors, appearing frequently.

Conducting separate analyses of each group for each site, followed by combining Near-NSs in one analysis and all NS speakers in another analysis, allows us to compare the influences on *ne*-retention across each group of speakers while identifying more global trends. No combination of factors was identical across groups, but each site shared several significant factors. Subject type, Verb type, Hiatus, and Emphasis were significant factor groups at each site for Near-NSs, while Subject type, Hiatus, Emphasis, and Interlocutor ID were significant factor groups at each site for NSs. The latter factor group, as mentioned earlier, shows high correlation with other factor groups, and is thus less likely to be a significant factor on its own. This is certainly due to the random variation that may occur when these speakers are sampled across many different conversations and interlocutors, but it can be instructive to see which Near-NS participants facilitate higher versus lower *ne*-retention in their interlocutors' rates. For example, speaker 5L produced the highest *ne*-retention in Lille Near-NSs; his performance may have influenced higher retention rates in his NS interlocutor compared to a randomly sampled speaker.

In the following subsections, I provide cross-tabulations on specific factor groups that influenced *ne*-retention in the variationist analyses, with additional commentary and comparison with previous studies where appropriate. I will then conclude with an overall summary of the conclusions drawn from the variationist analyses for these Near-NSs and their interlocutors.

4.5.9.1 Subject type

Subject type consistently appeared among the most influential variables across all groups of Near-NS speakers and interlocutors. As may be expected based on previous research, NPs and clauses that did not contain a subject favored *ne*-retention.

Ashby (1976), Coveney (2002), Diller (1983), Hanson and Malderez (2004), and Meisner (2016) all found that the relative pronoun *qui* also favored *ne*-retention. Though they did not conduct variationist analyses with *qui* as a separate factor from other subjects, I ran initial variationist analyses with *qui* as a fourth factor in the Subject type group. These analyses, however, consistently indicated that *qui* actually *disfavored* retention, which did not concord with the raw percentages of retention with *qui* (44.8% retention compared to 13.2% overall). In these analyses, Rbrul also indicated that Subject type had high collinearity with another factor for certain speaker groups, an issue that I will discuss in the following subsections. Table 4-30 provides the overall breakdown of *ne*-retention by subject type across all speakers (here, Near-NSs are grouped together across both sites; NS interlocutors are likewise all grouped together; in the rest of these subsections, "NNSs" refers to the two near-native interlocutors in Lille).

Table 4-30. Ne-retention by subject type

Subject type	**Near-NSs**	**NSs**	**NNSs**	**Total**	**% *ne*-retention**
qui	30/61	23/60	3/4	56/125	44.8%
none[57]	18/27	7/21	2/6	27/54	50.0%
NP	55/77	18/45	6/7	79/129	61.2%
other pronoun	278/1896	100/1796	30/304	413/3996	10.3%

[57] For Subject type = none, the majority of tokens were of negated infinitives, e.g., (*j'ai pris la decision de ne pas me spécialiser* 'I made the decision to not specialize'). Of the remaining tokens, 10 were imperatives without an expressed subject. Two were reported/quoted speech of a hypothetical teacher giving a command to a student (*Réponds pas; ouvre pas la bouche*, 'Don't answer; don't open your mouth'); of the remaining five imperatives, four were the expression *N'en parlons pas* ('Let's not talk about it').

When separating *qui* from other subject pronouns, we see that *qui* favors *ne*-retention at nearly the same rate as non-subjects and NPs—all much higher than the *ne*-retention found in subject clitics.

Regarding individual native speakers, much of the *ne*-retention results may be explained via the single linguistic factor of Subject type. Table 4-31 breaks down *ne*-retention by subject type for the three bilingual interlocutors in Pau.

Table 4-31. Ne-retention by subject type for bilingual interlocutors in Pau

	Ch		**Fr**		**Th**	
Subject type	*ne* / total	% *ne*-retention	*ne* / total	% *ne*-retention	*ne* / total	% *ne*-retention
Lexical subjects	8/15	53.3	2/16	12.5	2/3	66.7
qui	3/12	25.0	4/15	26.7	9/9	100.0
Subject pronouns	12/687	1.7	3/399	0.8	38/127	29.9
none	0/0	--	2/5	40.0	0/1	0.0
Total	23/714	3.2	11/435	2.5	49/140	35.0

Nearly half of all *ne* tokens (n = 11) for Ch were produced in the 27 tokens with lexical subjects and the relative pronoun *qui*. For Fr, the distribution is even more striking, with all but three *ne* tokens coming in the 36 negation contexts of lexical subjects, *qui*, and non-subjects. It is also noteworthy that of all native speakers, Fr has the lowest *ne*-retention with lexical subjects. This result is not necessarily due to repetition of a specific noun, as these 16 tokens contain 14 different lexical subjects. It simply seems that Fr's overall strong tendency to exclude *ne* is reflected in negation behavior with lexical subjects as well, though small token counts certainly preclude definitive conclusions. In any case, Fr very nearly has a pattern of complementary distribution, in which *ne* is deleted if the subject is a clitic pronoun but may be retained if the subject is anything else. Moreover, of Fr's 16 tokens with *qui*, four were in hiatus contexts, with two of these retaining *ne*. As for bilingual Th, his *ne*-retention by subject type has a similar pattern despite a smaller

sample size and comparatively higher *ne*-retention overall. Th categorically retained *ne* with relative pronoun *qui* as subject. Four of the nine instances of *qui* for bilingual Th are in hiatus contexts, with the other five in non-hiatus contexts. For this speaker, since all negation tokens with *qui* have *ne*-retention, and since all six hiatus tokens (regardless of subject) have *ne*-retention, it is impossible to determine to what extent the retention of *ne* is due to *qui* or to hiatus. Clearly, however, the interaction of *qui* as relative pronoun and *qui* introducing a hiatus context has a significant role in the appearance of *ne* overall.

4.5.9.2 Hiatus

Whereas the phonological environment for *ne*-retention has been the subject of some contention in previous studies (cf. section 4.5.2), reducing the effect of phonological environment to hiatus contexts in the current study produced a significant factor group for all Near-NS groups and interlocutors, as well as for SA learners. However, these results warrant a continuation of the discussion concerning the pronoun *qui.* Since *qui* contains a final vowel, a following verb with a vowel onset can produce a hiatus (e.g., *une personne qui est sympa*, 'a person who is nice'). Sometimes speakers can elide the /i/ in *qui* in these contexts (e.g., *une personne qu'est sympa*), but a careful examination revealed no elision of this type in the current study. In any case, almost half of the hiatus contexts produced by Near-NSs and interlocutors were due to relative pronoun *qui*, as Table 4-32 shows.

Table 4-32. Ne-retention in hiatus contexts: qui *versus other subjects*

Hiatus contexts	**Near-NSs: Pau**	**Near-NSs: Lille**	**NSs: Pau**	**NSs: Lille**	**NNSs**	**Total**	**% *ne*-retention**
qui	17/25	6/9	8/11	6/10	2/2	39/57	68.4
other	28/32	12/13	6/17	6/7	3/4	55/73	75.3
Total	45/57	18/22	14/28	12/17	5/6	94/130	72.3

Ne-retention percentages are only slightly lower for hiatus contexts involving *qui* compared with hiatus involving all other subjects, and this difference is not significant ($\chi^2(1) = .766$; $p = .382$). Thus, the hiatus following *qui* does not meaningfully favor or disfavor *ne*-retention more than hiatus following other subjects. Breaking down the subjects in the "other" category, we see in Table 4-33 that subject pronouns (mainly consisting of *ça*) and NPs alike strongly favor *ne*-retention in hiatus contexts compared with overall retention (13.2%).

Table 4-33. Ne-retention in hiatus contexts for other subjects

Hiatus contexts	**Near-NSs: Pau**	**Near-NSs: Lille**	**NSs: Pau**	**NSs: Lille**	**NNSs**	**Total**	**% *ne*-retention**
subject pronouns	14/18	11/12	3/13	4/5	2/3	34/51	66.7
NPs	14/14	1/1	4/5	2/2	1/1	22/23	95.7

We furthermore see that NPs in hiatus contexts almost categorically favor *ne*-retention (with the sole deletion produced by bilingual Fr: *beaucoup de petits ont pas lu nécessairement* 'a lot of the little ones haven't necessarily read'). All but three of the 51 subject pronouns involved *ça* followed by a vowel-initial verb; Near-NSs almost categorically retained *ne* in these contexts (e.g., *ça n'a rien à voir* 'that has nothing to do with it') while NSs were more likely to delete *ne* and allow the hiatus (e.g., *ça existe pas* 'that doesn't exist'). In general, however, it is clear that hiatus contexts favor *ne*-retention, across different subject types, and not solely due to the presence of *qui*.

Recall that in an initial variationist analysis, *qui* appeared to disfavor *ne* even though its presence resulted in higher retention than overall averages. This appears to be due, at least in part, to the interaction between *qui* and hiatus. Table 4-34 provides the results for all instances of subject pronoun *qui*, in hiatus and non-hiatus contexts.

Table 4-34. Ne-retention with qui *in hiatus and non-hiatus contexts*

***Qui* contexts**	**Near-NSs: Pau**	**Near-NSs: Lille**	**NSs: Pau**	**NSs: Lille**	**NNSs**	**Total**	**% *ne*-retention**
hiatus	17/25	6/9	8/11	6/10	2/2	39/57	68.4
non-hiatus	5/19	2/6	8/27	1/14	1/2	17/68	25.0
Total	22/44	8/17	16/36	7/24	3/4	56/125	44.8

We see that non-hiatus *qui* favors retention only slightly higher than the overall 13.2% retention rate. Thus, results from previous studies (cf. Ashby, 1976; Coveney, 2002; Meisner, 2016) showing *qui* to favor retention may in fact be reflecting the hiatus context more so than other characteristics of *qui*.

4.5.9.3 Subordinate clauses

Based on previous studies (Ashby, 1976; Donaldson, 2017; Meisner, 2016; Regan, 1996), I expected that subordinate clauses would favor *ne*-retention. Table 4-35 shows that this was the case for all speaker groups.

Table 4-35. Ne-retention in subordinate clauses

	Near-NSs: Pau	**Near-NSs: Lille**	**NSs: Pau**	**NSs: Lille**	**NNSs**	**Total**
% *ne*: main clauses	18.3 (188/1029)	14.1 (104/735)	4.7 (52/1111)	8.7 (48/551)	11.9 (34/286)	11.5 (426/3712)
% *ne*: subordinate clauses	32.8 (64/195)	24.5 (25/102)	17.4 (31/178)	20.5 (17/83)	20.0 (7/35)	24.3 (144/593)
% *ne* overall	20.6	15.4	6.4	10.3	12.8	13.2

Even though the variationist analysis selected Clause type only for the NSs in Pau and NSs overall, all speaker groups indeed produced *ne*-retention in subordinate clauses at higher rates than in main clauses. Note that this result also obtained for SA interlocutors (12% retention in 25 subordinate clauses versus 6.7% in the remaining 285 main clauses).

The type of subordinate clause also conditions *ne*-retention differently, as Table 4-36 shows.

Table 4-36. Ne-retention by type of subordinate clause

Subordinate clause	**Near-NSs: Pau**	**Near-NSs: Lille**	**NSs: Pau**	**NSs: Lille**	**NNSs**	**Total**	**% *ne***
complement	17/50	4/24	5/33	5/21	2/10	33/138	23.9%
relative	24/56	12/32	17/43	11/30	3/11	67/172	39.0%
adverbial	23/89	9/46	9/102	1/32	2/14	44/283	15.5%
Total	64/195	25/102	31/178	17/83	7/35	144/593	
% *ne*	32.8%	24.5%	17.4%	20.5%	27.2%	24.3%	

Ne-retention in adverbial clauses is only slightly higher than the overall retention rate, whereas complement clauses more strongly favor retention; relative clauses favor retention the most, and this is consistent across all speaker groups.

When breaking down relative clauses by type of relative pronoun, we also see a consistent pattern: *où* inhibits *ne*-retention while *qui* most strongly favors *ne*-retention (Table 4-37).

Table 4-37. Ne-retention with relative pronouns preceding negation contexts

Relative pronoun	**Near-NSs: Pau**	**Near-NSs: Lille**	**NSs: Pau**	**NSs: Lille**	**NNSs**	**Total**	**% *ne***
qui	22/44	8/17	16/36	7/24	3/4	56/125	44.8%
que	2/9	4/11	1/4	4/4	0/4	11/32	34.4%
où	0/3	0/3	0/3	0/2	0/3	0/14	0.0%
dont		0/1				0/1	0.0%
Total	24/56	12/32	17/43	11/30	3/11	67/172	
% *ne*	42.9%	37.5%	39.5%	36.7%	27.2%	39.0%	

Again, part of this favoring effect of *qui* may be due to the interaction of *qui* and hiatus. If we exclude *qui* in hiatus contexts, the *ne*-retention totals in Table 4-38 are produced for the remaining relative pronouns introducing clauses with verbal negation (along with *ne* percentages for these relative clauses compared with overall *ne* percentages for each group).

Table 4-38. Ne-retention in non-hiatus relative clauses

	Near-NSs: Pau	**Near-NSs: Lille**	**NSs: Pau**	**NSs: Lille**	**NNSs**	**Total**
relative pronouns (non-hiatus)	6/30	6/23	9/34	5/20	1/9	27/116
% *ne*	20.0	26.1	26.5	25.0	11.1	23.3
% *ne* overall	20.6	15.4	6.4	10.3	12.8	13.2

Thus, relative pronouns indeed slightly favor *ne*-retention in comparison to non-relative clause contexts, though this effect is not produced in all speaker groups, and the sample size is rather small. Nevertheless, we see that hiatus contexts in general favor *ne*-retention much more strongly than when relative pronouns in general introduce clauses with verbal negation.

4.5.9.4 Reinforcing adverb

The presence of a reinforcing adverb had a mixed effect on *ne*-retention; it was selected as a significant factor only for Near-NSs in Lille, where it in fact slightly inhibited *ne*-retention, while for other groups a reinforcing adverb slightly favored retention. As we see in Table 4-39, there are also some differences depending on the type of adverb involved.

Table 4-39. Ne-retention with reinforcing adverbs in negation contexts

adverb	**Near-NSs: Pau**		**Near-NSs: Lille**		**NSs: Pau**[58]		**NSs: Lille**		**NNSs: Lille**		**Total**	**% *ne***
du tout	8/26	30.8	1/32	3.1	3/35	8.6	1/43	2.3	1/7	14.3	14/143	9.8
même	4/23	17.4	0/10	0.0	1/19	5.3	1/14	7.1	0/2	0.0	6/68	8.8
surtout			1/2	50.0	0/2	0.0					1/4	25.0
-ment	4/14	28.6	4/9	44.4	4/28	14.3	0/4	0.0	1/1	100.0	13/56	23.2
Total	16/63	25.4	6/53	11.3	8/81	9.9	2/60	3.3	2/10	20.0	34/271	12.5
Overall % ne		20.6		15.4		6.4		10.3		12.8		13.2

[58] One utterance from this group (speaker Ch) contained both a *–ment* adverb and *du tout* (*c'est vraiment pas donné du tout* 'it's really not a given at all'). This utterance was thus counted for the *–ment* totals and the *du tout* totals, but not counted twice for the total adverb contexts.

Given the global *ne*-retention rate across all Near-NSs and interlocutors at 13.2%, it appears that reinforcing adverbs have little effect on *ne* overall. However, adverbs specifically ending in *–ment* seem to slightly favor *ne*-retention, at 23.2%, and it remains to be seen whether *surtout* would produce a similar favoring effect in a larger sample.

4.5.9.5 Emphasis

Turning now to sociostylistic factors, Emphasis was the most common significant sociostylistic factor group, appearing in all speaker groupings for Near-NSs and their interlocutors. Table 4-40 breaks down the emphatic negation totals for each speaker group.

Table 4-40. Ne-retention in emphatic negation contexts

Speaker group	***ne* / total: emphasis**	**% *ne*-retention: emphasis**	**overall % *ne*-retention**	**% increase for emphatic negation**
Near-NS Pau	35/59	59.3	20.6	38.7
Near-NS Lille	21/39	53.8	15.4	38.4
NS Pau	16/58	27.6	6.4	21.2
NS Lille	13/32	40.6	10.3	30.3
NNS Lille	8/16	50.0	12.8	37.2
Total	93/204	45.6	13.2	32.4

As elaborated by Fonseca-Greber (2007), emphatic negation can be observed in a variety of contexts including lexical emphasis, which may consist of a reinforcing adverb. Thus, there may be some overlap between these two factor groups. Unlike reinforcing adverbs, however, emphasis in general produces consistently high *ne*-retention rates across speaker groups. The most common type of emphasis observed was pitch prominence, usually occurring on the post-verbal negator (e.g., *mais je n'ai RIEN compris* 'but I understood NOTHING').

Individual speaker styles, and the topics of conversation, also condition the appearance of emphatic contexts. For example, speaker 4L produced 23 tokens of emphatic negation (with 69.6%

retention), reflecting his rather animated speaking style as well as an extended conversation topic about his strong desire to remain in his current professional teaching role rather than attempt promotion within the national education system. The remaining eight Lille Near-NSs, on the other hand, produced 16 emphatic negation tokens combined.

Finally, emphasis is one of the few contexts in which *ne*-retention is substantially more likely with frequent subject clitics such as *je* and *ce*. Expressions such as *c'est* / *c'était* and *je suis* / *j'ai* tend to resist *ne*-retention in the presence of linguistic factors such as a reinforcing adverb and negators other than *pas*, as well as other sociostylistic factors such as serious topics and formal quoted speech. However, when speakers draw attention to their speech in an emphatic context such as for contrastive effect, *ne*-retention is much more common (as demonstrated in the following utterance by speaker 7P: *la langue c'est pour communiquer…ce n'est PAS pour conjuguer correctement les verbes irréguliers* 'language is for communicating…it is NOT for conjugating irregular verbs correctly').

4.5.9.6 Quoted speech

Quoted speech approached significance as a factor favoring *ne*-retention in NSs in Lille, but the factor group was only significant for NSs in the all-site variationist analysis. Table 4-41 reports the descriptive results for quoted speech (*ne*-retention/total tokens).

Table 4-41. Ne-retention in quoted speech

	Near-NSs	**NSs: Pau**	**NSs: Lille**	**NNS**	**Total**	**% ne**
Formal context	6/23	2/8	3/11	3/9	14/51	27.5%
Neutral/informal	5/49	0/52	1/13	0/4	6/118	5.1%
Not quoted	367/1988	81/1229	61/610	38/308	547/4135	13.2%

As the above table indicates, all speaker groups often dropped *ne* even when quoting speech from putatively formal styles. This was borne out by contexts in which the speaker was quoting speech addressed to an interlocutor by polite *vous*, and in which *ne* was dropped. Such contexts occurred in native speaker utterances (e.g., *Monsieur le prof, vous avez pas cet accent* ('Teacher, sir, you don't have that accent')) and Near-NS utterances (e.g., *Il semble que vous faisez*[59] *pas assez d'heures*, 'It seems that you don't have enough hours'). On the other hand, quoting from a putatively neutral or informal style tended to decrease *ne*-retention. Most of these quotes concerned quotes to/from friends or family members, so it is unsurprising to find the lowest *ne*-retention in this category.

Determining the level of formality in quotes from teachers to students, and vice-versa, was somewhat problematic, as addressed briefly in section 4.5.6. When polite or informal pronouns are not used (neither in the negated clause nor in the surrounding context), it is not always obvious whether the quoted speech is directed toward a specific teacher or student, or whether the quoted speech is presumed to be simply reflecting the quoted speaker's thoughts. This can be illustrated in the following example, quoted by near-native SaE, referring to L2 English students struggling to read Shakespeare, *le langage, ils disent "je peux pas"* ('that kind of language, they say, "I can't"'). This example appears to be quoting the speakers' thoughts, and the surrounding context indicates that this quote is not a specific utterance from a specific student. I did not consider these examples to be formal.

Another, more general, issue concerns presumed quoted speech that may in fact be a summary of speech directed to another interlocutor, e.g., a conversation with a prospective employer speaking to a native speaker ('CaF'), *On peut pas vous prendre* ('We can't take (hire)

[59] The verb here is incorrectly inflected (correct: *vous faites*), which I interpret to be a speech error by the Near-NS, not by the person whom she quoted.

you'). This example is taken from a context that does not appear to be a specific quote, yet the style would clearly be formal, and was coded as such.

Despite these methodological considerations, quoted speech, specifically from a formal context, appears to follow the same pattern as other factors such as hiatus contexts: higher *ne*-retention in substantially smaller sample sizes compared to neutral, informal, or non-quoted speech. As in Donaldson (2017), this factor slightly favors *ne*-retention in Near-NS and NS groups alike.

4.5.9.7 Serious topics

Among the topics considered "serious" as outlined in previous literature (cf. section 4.5.6), some appear in the current corpus rather frequently, due to the fact that a majority of all speakers are (or were) teachers (with "education" considered a serious topic, and a rather broad one at that), as well as the fact that the primary NS interlocutor in Lille is a legal analyst, and so "the legal system" was a recurring topic in her speech. On the other hand, a frequent topic, given the timeframe of the Near-NS conversations in summer 2016, was the issue of the "Brexit" vote. Based on previous literature, I did not classify this topic as among the serious topics, though some speakers discussed aspects of this topic that I interpreted to fall under the topics "the legal system" and "moralizing." Finally, I did not include tokens of quoted speech in the topic analysis, since quoted speech can engender a "microlevel style shift" (cf. Donaldson 2017: 147) that can take precedence over the seriousness of the topic (or lack thereof). As Table 4-42 shows, serious topics increased overall *ne*-retention by about 5%, with each speaker group showing an increase between 1% and 6% for serious topics.

Table 4-42. Ne-retention in serious topic contexts

Speaker group	*ne* / total: serious topics	% *ne*-retention: serious topics	overall % *ne*-retention	% increase for serious topics
Near-NS Pau	93/366	25.4	20.6	4.8
Near-NS Lille	51/245	20.8	15.4	5.4
NS Pau	45/361	12.5	6.4	6.1
NS Lille	24/180	13.3	10.3	3.0
NNS Lille	15/110	13.6	12.8	0.8
Total	228/1262	18.1	13.2	4.9

It is important to note that there is likely some selection bias in play here. For example, one speaker, 4L, produced relatively high *ne*-retention overall compared to other Near-NSs, and his negation tokens contained a larger percentage of serious topic tokens compared with other Near-NSs. Furthermore, recall that bilingual Ch and Fr both had their highest retention rates with speaker 2P. Across these two conversations, 54 of their 88 negation tokens with this speaker were produced during serious topics, including all 8 *ne*-retention tokens. That is, 61% of their negation tokens with this speaker were produced during serious topics, compared with 26% of negation tokens produced during serious topics in all other Pau conversations. The longer portion of the conversation spent discussing these topics certainly seems to have favored the production of *ne*-retention. Based on these considerations, then, topic appears to play at least a small role in influencing *ne*-retention.

4.5.9.8 *Tu* versus *vous*

Before discussing the variationist analysis reporting of *tu* versus *vous,* some methodological considerations are necessary. When meeting with each participant to brief him/her on the speaking task, I explained that the task would be an informal conversation. I did not give any guidance on which pronouns of address (*tu* versus *vous*) to use, so each participant was required to negotiate these pronouns with each interlocutor. Often the participants did this

explicitly (e.g., *On se tutoie ?* 'We'll use *tu*?') as their first exchange to begin the conversation. Other times the first speaker simply began using *tu* and the interlocutor followed suit. Generally, dyads matched more closely in age chose *tu*, which may also explain the choice of *tu* for all dyads in Lille,[60] where participants were relatively young compared to Pau. As for Pau, all of the NSs were younger than their interlocutors by at least 10 years, and so due to politeness or caution, the NSs generally used *vous*, even after the Near-NS interlocutor had negotiated for *tu* and used *tu*. There was no evidence that the NS adoption of an English identity led to using *tu*. In one case (the 4P/ThE dyad), both speakers used *tu* in the first three minutes and there were no pronouns of address until about 11 minutes, when 4P began using *vous* and ThE followed suit for the remainder of the conversation. In this case, I coded the negation tokens for *tu* in the first 11 minutes and *vous* thereafter. Also note that there were several instances of using *tu* when not directed to the interlocutor, being used as a general pronoun of address (e.g., talking about choosing school subjects, *quand tu parles bien anglais tu vas pas en S*, 'when you speak English well, you're not put on the science track'). These tokens were not coded differently when *vous* was used as the pronoun of address toward the interlocutor.

As for the variationist analysis, the adoption of *tu* versus *vous* appeared to significantly influence *ne*-retention in NSs in Pau and in the combined analysis of NSs at both sites, though the influence is somewhat deceptive. It is possible that this factor led the NSs to subconsciously reflect on the formality of the situation, with choices made regarding *ne*-retention in line with what one

[60] In one conversation in Lille, the native speaker (CaF) used *vous* with the near-native interlocutor (SaE) recruited to serve as conversation partner with the Near-NSs, whereas SaE used *tu*. This conversation was not included in the data analysis (see section 4.4.3.1). Though both speakers were female and differed by fewer than 10 years in age, one explanation for CaF's use of *vous* may be because this was CaF's first conversation as the primary native speaker recruited for the project in Lille. CaF's *ne*-retention in this conversation was only slightly higher (15.0%; 9/60) than in the remaining conversations (10.9%; 58/532), suggesting that the adoption of *vous* (or CaF's adoption of a level of formality deemed appropriate for this particular interaction) had only a marginal effect on *ne*-retention at best.

would expect: lower *ne*-retention for those speakers who used *tu* compared to the speakers who used *vous*, when comparing NSs with those at the same site. However, since more of the Pau NSs used *vous* and these speakers had lower *ne*-retention overall compared to Lille NSs, the differences in both sites are levelled (overall: *tu*: 7.8%, 69/879; *vous*: 7.6%, 79/1044). Indeed, Pau NSs using *vous* had 7.6% retention while Lille NSs, who used *tu* in all conversations, had 10.3% retention.

The difference identified as significant in the variationist analysis for Pau NSs may also be partly accounted for by Th's conversation with speaker 1P at 63.8% *ne*-retention while using *vous*; no other bilingual produced higher than 21.7% *ne*-retention in a single conversation. For NSs in Pau, Table 4-43 breaks down the *tu/vous* differences by interlocutor.

Table 4-43. Ne-retention in Pau NSs: tu *versus* vous

	When using *tu* with interlocutor		**When using *vous* with interlocutor**	
Speaker ID	*ne* / total contexts	% *ne*-retention	*ne* / total contexts	% *ne*-retention
Ch	2/109	1.8	21/605	3.5
Fr	2/133	1.5	9/302	3.0
Th	0/3	0.0	49/137	35.8
Total	4/245	1.6	79/1044	7.6

Excluding the outlier in Th's conversations reduces the *ne*-retention rate for *vous* to 4.9%. Though the *ne*-retention increase while using *vous* is minimal for Ch and Fr, it is nevertheless possible that this choice of address pronoun influenced *ne*-retention to a small extent.

4.5.9.9 Conclusions for Near-NSs

The variationist analyses reported in this section largely support previous findings concerning the factors that affect *ne*-retention. Some factors overlap across both L1 and L2 French speakers, supporting the argument that Near-NSs are conditioned by the same factors as NSs. None

of the factors showed trends in the opposite direction of those identified in previous literature (with the possible exception of conversation portion, obtained from the small sample of Lille Near-NS interlocutors). These findings thus broadly support evidence that these factors have been relatively stable on a diachronic level, even given the lower overall use of *ne* in more recent surveys. Furthermore, the language background of the interlocutor, significant in SA learners, was significant in none of the Near-NS groupings. This interlocutor effect was presumed to be insignificant based on the overall *ne*-retention percentages across interlocutor types in the Near-NS groups; however, the variationist analysis of these Near-NSs supports the hypothesis that with increased proficiency, the interlocutor effect diminishes in influence. The adoption of an English or French identity by bilingual interlocutors likewise did not significantly impact *ne*-retention in Near-NSs.

4.6 Hypotheses revisited: *Ne*-retention

I now revisit the research questions and hypotheses outlined in Chapter 3, as they apply to *ne*-retention. The first research question concerns the sociolinguistic ability demonstrated in learners of French at various proficiency levels; the hypothesis was that intermediate to advanced learners would retain *ne* at rates closer to standard (or classroom) norms than near-native speakers, while near-native speakers would demonstrate at least some overlap with NS patterns. For *ne*-retention, this hypothesis was confirmed at a group level, concording with previous studies. Overall, SA learners produced significantly higher levels of *ne*-retention than both their near-native and NS interlocutors, though there was much individual variation. For the Near-NS group, *ne*-retention overall was much closer to NS production, with some overlap (in terms of raw percentages as well as in variationist analysis factors); as a group, however, Near-NSs still retained *ne* at significantly higher overall rates than in the NS interlocutor groups.

The second research question concerns the role of the interlocutor language background and its influence on sociolinguistic variation, with the hypothesis that there would be convergence between lower-level learners and their interlocutors; near-native speakers would demonstrate less sensitivity to this interlocutor effect, though targetlike *ne*-retention patterns may only be demonstrated in interactions with NS. For *ne*-retention, hypothesis #2 was partially confirmed. On one hand, SA learners did show evidence of convergence toward their interlocutors (as Dewaele (2004a) found), with higher *ne*-retention with a near-native and a NS interlocutor than with other SA learners; the variationist analysis confirmed the status of the interlocutor as a significant factor (and the only non-linguistic factor). On the other hand, the study of Near-NSs showed no significant effect of the interlocutor's language background, regardless of differences in perceived or actual interlocutor L1 identity. However, while the Near-NS data suggest that targetlike behavior in terms of *ne*-retention rates was not necessarily facilitated by speaking to NSs, the absence of a NS speaker in the sphere of conversation does not inhibit nativelike *ne*-retention patterns for some learners.

Chapter 5: Subject doubling

The second sociolinguistic variable analyzed in the current study is subject doubling. As the previous chapter has shown, there is evidence of interlocutor effects conditioning sociolinguistic variation in certain learner groups for the variable under question, namely, *ne*-retention. A crucial question, then, is whether such effects can be detected in the use of other morphosyntactic structures in French that vary according to the formality of a spoken interaction, that are likely to appear at quantitatively sufficient frequency in the current corpus of informal, spontaneous oral discourse, and whose detection falls under a sufficient threshold of ambiguity.

As introduced in Chapter 3, like *ne*-retention, potential contexts for subject doubling in French are expected to be highly frequent in conversational interactions of the type obtained in the current corpus, since any clause containing a lexical NP subject is a candidate for doubling (as section 5.1 will detail). Furthermore, like *ne*-retention, subject doubling can be considered a binary variable (presence or absence of a doubling clitic), making it favorable to a variationist analysis, and like *ne*-retention, it involves a superficially redundant grammatical form (from an L1 English-L2 French learner's perspective) that is conditioned by stylistic variation. Moreover, *ne*-retention and subject doubling occur in the same morphosyntactic "neighborhood"; it can be instructive to observe, as some studies have done (cf. Villeneuve & Auger, 2013), how these two variables interact. Specifically, does informal L2 French combine *ne*-deletion with the presence of subject doubling in ways that mirror what is observed in L1 French? Finally, unlike *ne*-retention, there is a dearth of studies specifically focusing on subject doubling in L2 French, especially at lower proficiency levels; the current study can thus significantly expand the previous scope of inquiry by including learners at multiple proficiency levels and by comparing the results from the current corpus to varieties of French analyzed in the few existing studies.

In this chapter, the first three sections contain a description of the characteristics of subject doubling in French and provide an account of previous literature on L1 and L2 French speakers. In section 5.4, I report the frequency and nature of subject doubling use for all groups of speakers in the current study. This is followed in section 5.5 by a description of the variationist analysis conducted on this variable using the data from my corpus. I then conclude the chapter by revisiting my initial hypotheses in light of the results obtained for this variable.

5.1 Background on subject doubling in French

Subject doubling (hereafter abbreviated SD) is the co-occurrence of a subject clitic and a subject strong pronoun or noun phrase. As (1) demonstrates, in colloquial or informal styles of French, an NP subject such as *ma mère* can be immediately followed by a subject clitic anaphor, *elle*, which attaches to the verb. In standard French, this clitic is absent.

(1) a. Standard French: *Ma mère arrive demain.* 'My mother arrives tomorrow.'
b. Colloquial French: *Ma mère elle arrive demain.* 'My mother (she) arrives tomorrow.'

Such a doubling structure is not unique to French; it is found in other Romance languages such as Italian (dialects in northern Italy; cf. Rizzi, 1986) but also in typologically more dissimilar language families such as Finnic (see Holmberg & Nikanne (2008) for Finnish) as well as Bantu (see Bresnan & Mchombo (1987) for Chichewa).

5.1.1 Strong pronouns and subject doubling

In both standard and colloquial French, strong pronouns can be doubled by a corresponding subject clitic (2b), and this phenomenon is widespread across all regional varieties of French. However, strong pronouns differ from lexical subjects in terms of whether they can function as subjects without co-occurring subject clitics (3).

(2) a. *J'arrive demain.* 'I arrive tomorrow.'
b. *Moi, j'arrive demain.* '(Me,) I arrive tomorrow.'

(3) a. *Moi, j'arrive demain.* '(Me,) I arrive tomorrow.'
b. **Moi arrive demain.*[61]
c. *Toi, tu arrives demain.* '(You,) you arrive tomorrow.'
d. **Toi arrives demain.*
e. *Lui, il arrive demain.* '(Him,) he arrives tomorrow.'
f. *Lui arrive demain.*

As these examples demonstrate, in Hexagonal French first- and second-person strong pronouns cannot occupy the subject position alone,[62] whereas third-person strong pronouns (*lui/elle/eux/elles*) can. The use of a strong pronoun in subject position alone, without a co-occurring subject clitic anaphor, as in (3f), is attested but much less common than with the subject clitic (as in (3e)), subject to pragmatic constraints and often denoting a contrastive function (4). By way of comparison, in English, the corresponding subject pronouns can receive emphatic stress, whereas subject clitics in French cannot receive emphatic stress (indicated by capital letters in (5)). This analysis is complicated somewhat by the fact that the strong subject pronouns *elle/elles* are identical to the subject clitics *elle/elles*; however, contrastive structures demonstrate the requirement for a strong subject pronoun (as in (6)) but an optional subject clitic.

(4) *Lui arrive demain alors qu'elle arrive samedi*.[63]
'*He* arrives tomorrow whereas *she* arrives Saturday.'

(5) **IL arrive demain alors qu'ELLE arrive samedi*.

(6) a. *Lui, il arrive demain alors qu'elle, elle arrive samedi.*
b. *Lui arrive demain alors qu'elle arrive samedi.*

[61] In Hexagonal French, Coveney (2003) specifies that one occasionally finds *moi seul/e* and *toi seul/e* as strong subject pronouns without co-occurring subject clitic, but only in written French.
[62] Strong pronouns in subject position without clitic anaphor (as in (3b)) are nevertheless possible in some regional varieties; cf. King & Nadasdi (1997: 277) for Acadian French in Newfoundland and Rottet (1995: 181) for Cajun French in Louisiana (in Terrebonne and Lafourche Parishes).
[63] To further emphasize this contrast, strong pronoun *elle* may be prosodically separated from surrounding segments via a lack of *liaison* and *enchaînement*, e.g., in (4) as [a.lɔʀ.kə.ɛl.a.ʀiv.sam.di], rendered orthographically as *…alors que elle arrive samedi*. In his study of spoken French, Carton (2009) found that two thirds of all 3SG/PL strong pronouns were prosodically separated from the preceding element.

Coveney (2003: 117) points out that sequences of 3SG/PL strong pronouns optionally doubled by a subject clitic only occur in contrastive contexts, which leads him to exclude these forms from his quantitative analysis of subject doubling. However, 3SG/PL strong pronouns can also have discursive functions, such as signaling a topic shift or an elaboration of topic (Stark, 1999). Moreover, the case of strong pronouns is interesting from a variationist (and acquisitional) perspective since, the question of contrast or topic aside, there can be variation in the *type* of doubling subject clitic co-occurring with the strong pronoun. This subject clitic can be a personal pronoun (*il/elle/ils/elles*) or a neuter demonstrative pronoun (*ce/ça*), demonstrated, for example, in *Lui il est le propriétaire / Lui c'est le propriétaire* ('Him he is the owner').

It should be noted, however, that while the masculine strong pronouns *lui* and *eux* are relatively easy to differentiate from either type of doubling clitic referent (*lui* with *il*; *eux* with *ils; lui* or *eux* with *ce/ça*), the feminine strong pronouns *elle/elles* are, at least for most speakers of Hexagonal French, identical in pronunciation to their personal clitic referents. In other regional varieties of French, *elle/elles* have different pronunciations depending on their syntactic function as strong pronoun or clitic; for example, in Saguenay (Québec) French (cf. Auger & Villeneuve, 2010), the strong pronoun form of *elle* is pronounced [ɛl], whereas its clitic form is typically pronounced [al] or [a]. On the other hand, with feminine strong pronoun doubling in Hexagonal French,[64] such as *Elle, elle est belle* ('Her, she is beautiful'), it may not always be possible to distinguish between a combination of strong pronoun and clitic or simply a repetition of the clitic (though pitch contours of the strong pronoun form of *elle* can serve to distinguish it from subject clitic *elle*). The complexity of *elle/elles* notwithstanding, variation is indeed possible with respect

[64] For some speakers of Hexagonal (non-meridional) French, in highly informal contexts, subject clitic *elle* may be realized as [al] or [a], as in Saguenay French (cf. Carton, 1987: 40). Though it is possible that learners may have interacted with NSs who produce such variants of *elle*, it is unlikely that these learners have encountered this variation in their input to such a degree that they have incorporated this variation into their own morphophonological inventory.

to certain strong pronoun subjects, as the third-person singular/plural forms can be optionally doubled by subject clitics.

5.1.2 The framework of subject doubling analyses

The use of the expression *co-occurrence* in this definition at the beginning of this section does not identify which element constitutes a "doubling" of the other element. That is, what is the underlying element to which an optional—or required—second, "doubling" element is added? In principle, it is possible to quantitatively analyze subject doubling from the perspective of the presence or absence of a subject NP (or strong pronoun) in clauses containing a subject clitic. For example, all four sentences in (7) would fall under the scope of investigation for possible subject doubling.

(7) a. *Je parle français.* 'I speak French.'
b. *Ils parlent français.* 'They speak French.'
c. *Moi, je parle français.* '(Me,) I speak French.'
d. *Les enfants, ils parlent français.* 'The children (they) speak French.'

Gadet (1997: 132) follows this methodology in her analysis of dislocation in spoken French, where she states, "On ne rencontre pas plus de 10% de sujets disloqués (certains corpus peuvent aller jusqu'à 15 ou 16%, jamais au-delà)."[65] Gadet's quantitative analysis, which uses data from Blanche-Benveniste (1994), includes *all* clause subjects, even those clauses consisting of a lone subject clitic pronoun (as in (7a) and (7b)), with the latter type representing more than half of all subjects in her data. It is certainly possible to frame the question in terms of how frequently subject NPs, or certain strong pronoun subjects, are dislocated from the verb, out of all clauses where co-occurring subjects are possible. However, Gadet's approach seems to be the exception regarding the framing in which quantitative analyses of subject doubling have been defined.

[65] "One does not find more than 10% of dislocated subjects (some corpora may go as high as 15 or 16%, but never higher)" (translation mine).

Nadasdi (1995a) addresses such framing of the subject doubling phenomenon when he points out that while certain strong pronouns in subject position (e.g., *moi*) require co-occurring subject clitics (in most varieties of French), the reverse is not true: a subject clitic does not require a co-occurring strong subject pronoun. That is, using my previous examples, ungrammatical (3b) requires a subject clitic (as in (3a)), but (2a) does not require a strong subject as in (3a).

Nadasdi subsequently motivates the approach of examining subject doubling in terms of the presence or absence of a subject clitic for reasons related to bound versus free morphemes and levels of language restriction. Though the latter discussions are beyond the scope of this study, the perspective from which subject doubling is analyzed is important for any treatment of this issue, whether theoretical or applied. Anticipating a variationist approach in the current study as Nadasdi did in his study, this framework establishes the dependent variable as the presence or absence of a subject clitic, rather than the presence or absence of a strong subject pronoun or subject NP. Though certain methodological considerations may differ (e.g., what constitutes subject doubling versus left-dislocation), other variationist studies on subject doubling have adopted the same basic framing of the question as outlined by Nadasdi (e.g., Coveney, 2003; Villeneuve & Auger, 2013; Zahler, 2014)—that is, all lexical NP subjects (and some strong pronoun subjects, depending on the study) are selected, and those that co-occur with subject clitic co-referents constitute subject doubling. Furthermore, Coveney's (2003) variationist analysis emphasizes that, in a treatment of variable SD, one must adopt an approach obtaining relative frequencies of the number of occurrences out of the number of total potential occurrences. Therefore, utterances consisting of a lone subject clitic were excluded from SD data analysis in the aforementioned studies. The current study will adopt this framework as well.

5.1.3 Subject doubling or left-dislocation?

In framing the scope of what constitutes subject doubling, a similar question arises regarding differences between subject doubling and left-dislocation. Both terms have been used to describe the sequence 'NP + subject clitic', though left-dislocation is a broader phenomenon in which the first element may consist of a syntactic phrase other than an NP; however, left-dislocation of an NP results in slightly different syntax than subject doubling. In this section, I outline the differences involved in these two structures and advance an argument restricting occurrences of subject doubling analyzed in my corpus to sequences of NPs followed by a clear co-referential subject clitic, based on previous studies of L1 and L2 French.

Many previous studies consider utterances such as (6a) (*Lui, il arrive demain alors qu'elle, elle arrive samedi*) in which doubling serves a contrastive function, to be examples of left-dislocation (Ashby, 1988; Barnes, 1985; Lambrecht, 2001). Left-dislocation (abbreviated hereafter as LD) refers to utterances in which a constituent has been moved from (or is not generated in) canonical subject, object, or other argument position and appears at the beginning (i.e., the left-periphery) of an independently grammatical clause, coindexed with a resumptive clitic in the clause; the constituent's appearance outside of canonical position serves to distinguish this constituent in some way, such as to establish a topic, change focus, add emphasis, or introduce contrast. Following Chomsky (1995), some scholars (Cinque, 1990; Kayne, 1994) consider dislocation to involve movement, while others (De Cat, 2007) argue that dislocated elements are generated where they surface. The dislocated element can also appear at the *end* of an independently grammatical phrase, known as right-dislocation, with a similar discourse function of establishing topic, focus, or emphatic contrast, though such structures will not be treated in the current study (see, e.g., Donaldson (2011b) for discussions therein). The most common constituent

appearing at the left periphery is an NP, with dislocated tensed clauses, infinitival phrases, prepositional phrases, and adjectival phrases appearing much less frequently (Donaldson, 2008).

When the dislocated element is the subject, left-dislocation may superficially appear as identical to subject doubling. However, structurally this results in slightly different syntax between SD and LD, as demonstrated by Figure 5-1 (Culbertson, 2010; Roberge, 1990).

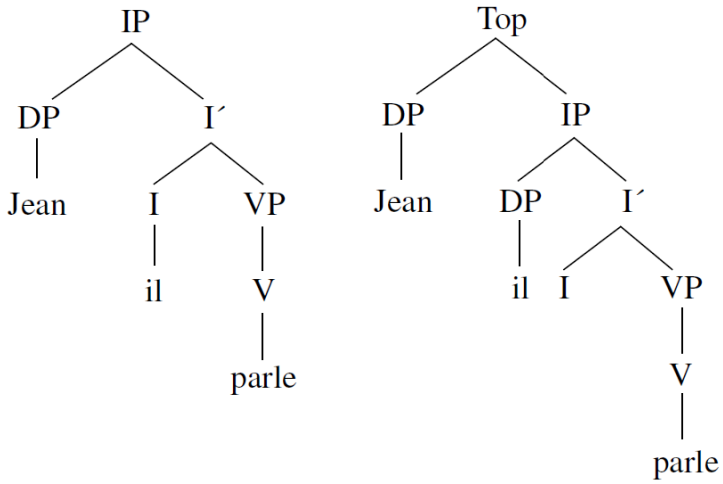

Figure 5-1. Syntactic structure of left-dislocation (left) versus subject doubling (right), from Culbertson (2010: 105)

In LD, the subject clitic is in canonical subject position with the DP outside the clause. In SD, the lexical noun occupies the subject position with the clitic reanalyzed as a verbal agreement marker with the DP (Auger, 1994). Note that other literature focusing on the discourse properties of French LD (e.g., Ashby, 1982; Barnes, 1985; Lambrecht, 1981, 1987, 1994, 2001) does not always distinguish between LD and SD, as Donaldson (2008) points out. Donaldson's study also does not specifically distinguish between LD and SD, subsuming under LD what other researchers would consider SD.

Whether the subject clitic is a true subject argument in canonical subject position or a verbal agreement marker cliticizing to the verb remains an outstanding debate in Romance

languages. In earlier work, according to the Clitic Hypothesis (cf. Kayne, 1975, 1991; Rizzi, 1986), it was posited that subject clitics were argument-bearing subjects in canonical subject position in the syntax (receiving case and a theta-role), but that these subjects cliticize to the verb at the post-syntactic surface level (in phonology). This argument has been supported in more recent work by Laenzlinger (1998), Belletti (1999) and De Cat (2005, 2007). While in some Northern Italian dialects, subject clitics appear to be obligatory agreement markers (Rizzi, 1986), in French a resumptive subject clitic is optional for all but a small number of NP subjects. De Cat (2005: 1217) further argues that a morphological analysis of French subject clitics "places a heavy burden on the lexicon" and that the distribution of elements intervening between a subject clitic and verb stem is syntactically constrained. However, the double occurrence of lexical (or strong pronoun) subjects and resumptive subject clitics is so frequent in colloquial French (encompassing multiple dialects) that many other scholars (e.g., Auger, 1994; Culbertson, 2010; Roberge, 1990; Zribi-Hertz, 1994) question whether French subject clitics are still syntactic subjects and suggest that they have been reanalyzed as verbal agreement markers (inflectional affixes). Culbertson represents the syntactic structure of subject clitics according to both hypotheses, as demonstrated in Figure 5-2.

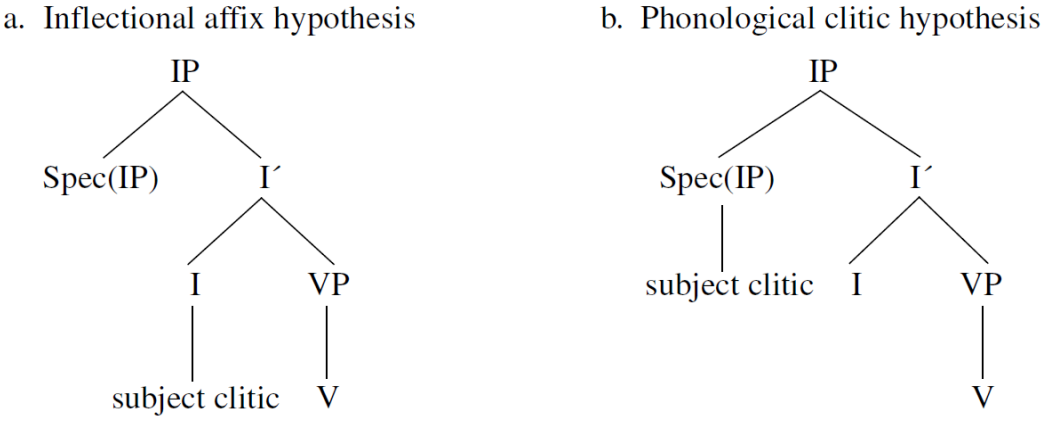

Figure 5-2. Syntactic structure of French clitics as verbal agreement markers (a) and as canonical subjects (b), according to Culbertson (2010: 89)

Culbertson (2010) provides three supporting arguments in favor of the structure in Figure 5-2a. First, despite evidence seemingly to the contrary (e.g., with the negative particle *ne* and object clitics; cf. De Cat, 2005), no nonaffixal material can intervene between the subject clitic and the verb, nor do subject clitics take wide scope over conjoined VPs, which would be expected only if subject clitics are inflectional affixes. Second, prosodic differences show that French subject clitics as true subject arguments would necessarily involve LD with a different prosodic contour than what happens in cases of subject doubling, where prosody is similar (or even identical) between monoclausal lexical subjects co-occurring with a subject clitic and lexical subjects that are not doubled. Third, native speaker acceptability judgments on SD in informal contexts, as well as data from L1 French children and child-directed speech, suggest that L1 acquisition encourages child learners to treat subject clitics as morphological markers of agreement. Further work by Palasis (2015) on child acquisition of clitics has strengthened the argument for analysis as agreement markers.

A counter-argument against the clitic-as-verbal-agreement-marker position is that if subject clitics are agreement markers, their use should not be subject to variation, since grammatical agreement is presumed to be categorical. Auger (1998), however, notes that such a position stems from a theoretical (and historically, an often prescriptive) perspective in which two forms cannot have precisely the same meaning or function in a given context—the alternations must come either from codeswitching or two grammars that are in competition. Auger provides evidence from Québec French that in informal styles, variation is possible within a single grammar (and presence or absence of a doubled subject clitic is intended by the speaker, rather than a performance error). Auger (2003) admits that consensus on this debate remains elusive, and

Donaldson (2008) contends that more corpus analysis of spoken French is needed in order to strengthen subsequent arguments on this matter.

The current study does not aim to resolve this debate nor provide supporting evidence for one argument over another concerning the nature of French subject clitics; likewise, it does not aim to redefine what constitutes LD versus SD. My analysis will focus more narrowly on the frequency of the sequence 'NP + subject clitic' and compare its usage in my corpus of L1 French and L2 French speakers, in order to determine the extent to which learners have acquired the grammar of colloquial French and this particular sociolinguistic norm. Nevertheless, a discussion of the specific types of doubling to be included in my analysis warrants further elaboration.

Structural considerations aside, Roberge (1990) and Nadasdi (1995a) point out that a distinction between LD and SD can often be made prosodically: LD occurs when there is a pause, emphatic or contrastive stress, or absence of consonantal *enchaînement* or *liaison* between the subject NP and a doubling clitic pronoun. However, the determination of such a pause, stress marker, or lack of *liaison* is not necessarily a straightforward indicator of dislocation. For example, Deshaies, Guilbault, and Paradis (1993) found that most left-dislocations in Québec French do not have a pause after the lexical noun, despite evidence of emphatic stress. In his study on SD, Nadasdi (1995a) motivates an approach for determining LD with two criteria: a lack of *liaison* and the presence of emphatic stress on the last syllable of the NP. This determination therefore relies on access to the audio recording of the utterance in question; conversely, analyses relying on written or transcribed corpora may not have access to detailed acoustic phenomena that allow for such determinations to be made. Some transcriptions of spoken French, and those provided in Nadasdi's corpus, for example, indicate a pause and/or lack of *liaison* (that is, left-dislocation) by

inserting a comma between the dislocated NP and the verbal argument. Nagy et al. (2003), quoting Auger (1995), provide another example in (8):

(8) a. Left dislocation: *Les maringouins, ils me suivent.* 'Mosquitos, they follow me.'
b. Subject doubling: *Les maringouins ils me suivent.* 'Mosquitos follow me.'

Using a comma to indicate a prosodic boundary, and thus indicate LD, is certainly a simple and easy transcription convention. However, this does not necessarily mean that a *lack* of comma indicates a *lack* of prosodic boundary. (Indeed, in the above example of SD, there can be no liaison between *maringouins* and *ils*.) Likewise, Dupont (1985: 68) found that when written sentences of the type in (9) were read, no difference in *enchaînement* was made between (9a) and (9b), as Dupont's transcriptions indicate.

(9) a. *Frédéric, il est venu* [fʀedeʀikəilɛvəny] 'Frédéric, he came.'
b. *Frédéric il est venu* [fʀedeʀikəilɛvəny] 'Frédéric (he) came.'
c. *Frédéric est venu* [fʀedeʀikɛvəny] 'Frédéric came.'

Non-dislocated elements, as in (9c), have obligatory *enchaînement*.

Acoustic analyses of dislocation likewise reveal less-than-straightforward prosodic differences than those described above (e.g., Deshaies et al., 1993; Nadasdi, 1995a). For example, Guilbault (1995) conducted multiple prosodic analyses of LD structures in recorded spontaneous speech. Of 131 dislocations, only 16 were determined to contain a pause. Furthermore, of 30 dislocations which could have resulted in *enchaînement*, only 6 were resyllabified to produce *enchaînement*, all from the same structure *eux autres ils* 'those others they' (where the [t] in [œ.zot(ʀ)] was resyllabified to [œ.zo.tʀi(l)]). Guilbault also identified 22 utterances in which the same speaker produced LD and non-LD with multiple tokens of the same lexical NP subject (e.g., *ma mère elle est encore en pleine santé* 'my mother (she) is still in good health' versus *ben là sa mère est morte il y a deux ans* 'well his mother died two years ago'; cited in Avanzi, 2012: 134). While native speakers in a perception task could reliably detect a stress marker on the final syllable

of dislocated elements, stress markers on non-doubled subjects were also detected at a frequency that did not significantly differ from doubled subjects. Thus, presence or absence of stress was not significant in determining dislocation. Avanzi (2012: 179) found similar results in an acoustic analysis of syllable prominence for doubled and non-doubled subjects; likewise, an analysis of *enchaînement* could not distinguish prosodically between doubled and non-doubled subjects.

Despite these considerations, motivating a difference between LD and SD when no other influences on prosody are in play (Carroll, 1982; Nadasdi, 1995a; Roberge, 1990) has generally not been problematic for these researchers when examining spoken data from native speakers. Concerning spoken data from L2 French speakers, however, other factors must be considered. Nagy et al. (2003) argue that it can be difficult to determine whether L2 French learners are in fact producing NPs in topic position, and that instances of pausing and lack of *liaison* may be due to delays in lexical retrieval or lack of fluency more generally, while emphasis on the NP may be due to transfer from L1 intonation patterns. In their study of L2 speakers of Montréal French, Nagy et al. did not distinguish between LD and SD, citing arguments in Deshaies et al. (1993) and Carroll (1982).

From a variationist standpoint, Coveney (2003) argues that Nadasdi's decision to exclude LD from his analysis of SD supposes that LD is not subject to variation; that is, a coreferential subject clitic is required, rather than optional, with all left-dislocated subjects, since the clitic functions as an agreement marker on the verb. Coveney provides, from his corpus, examples of stressed subject NPs followed by a pause but without a coreferential subject clitic, as in (10) (number 5 in Coveney's examples).

(10) *une maman* [pause] *va organiser les repas de la monitrice*. (p. 114)
'A mom [pause] is going to organize the instructor's meals.'

Coveney therefore concludes that differences between LD and SD based on prosodic determinations are not pertinent for his data analysis. For his native speaker data, Coveney (2003) adopts a classification in which SD includes all sequences of a full NP followed by a co-referential subject pronoun. While this includes some instances of what Nadasdi and others would call LD, involving topic shift or contrast, Coveney reasons that certain instances of LD still allow for a non-doubled structure to be used in the same context with the same value, providing examples where subject NPs can receive topic focus, such as when the speaker pauses or emphasizes the NP through pitch contrasts. On the contrary, contexts where the doubling pronoun is not co-referential with the subject NP—which, as Coveney stresses, may not be a straightforward judgment—do not constitute instances of SD. One of Coveney's examples (in (11) below) involves a context where a female speaker is talking about the difficulties in continuing with her career while taking care of two small children (example 14 in Coveney (2005); cf. examples 38 and 41 in Coveney (2003: 120)):

(11) *j'ai pas redemandé d poste—parce qu'avec les deux enfants s / avec un enfant ça va mais deux enfants c'est plus difficile.* (p. 102)
'I didn't ask for another [teaching] position—because with two children…with one child it's ok but two children it's more difficult.'

The NP in this example, *deux enfants*, is understood to be a topic, rather than the subject of the following verb (*c'est*). Here, a brief detour is warranted concerning the copula structure *c'est* and neuter demonstrative pronouns *ce/ça* more generally. When the NP is followed by a subject clitic, regardless of whether the NP is a subject or a left-dislocation, the clitic is what determines verbal agreement. In most cases, the verbal agreement would be the same, regardless of which subject determines agreement, as illustrated by the identical verb form *est* in (12).

(12) a. *Ma voiture elle est verte.* 'My car (she/it) is green.'
b. *Ma voiture est verte.* 'My car is green.'
c. *Elle est verte.* 'It is green.'

While this verbal agreement concording with the NP applies with the personal subject clitic pronouns *il/elle/ils/elles*, and while the choice of personal doubling pronoun is fairly straightforward when concording with the NP in gender and number agreement, the neuter demonstrative pronouns *ce/ça* behave somewhat differently. In Hexagonal French, *ce* (and its elided form *c'*) and *ça* have been analyzed as allophonic variants with phonologically conditioned allomorphy (e.g., Thibault, 1983; Truby, 1995), and while certain structures in informal French may allow either variant (where *ça* is the more colloquial form; cf. Ball, 2000: 70), the following verb, as well as style, condition the choice of variant. In spoken French, *ce* is used only with simple forms of *être* (e.g., present indicative/subjunctive/conditional/future), while *ça* may appear with both simple and compound forms (e.g., *passé composé*/past conditional) of *être*,[66] as well as with all other verbs (for an overview, see Piron, 2017: 121). Furthermore, *ce* and *ça* may appear as informal variants of *cela*, though there are constructions in which formal *cela* may not be interchangeable with *ce/ça* and vice-versa; note also that since *cela* is generally reserved for more formal styles, it would likely not co-occur with doubled NP subjects (cf. Truby, 1995: 36). In addition, *ça* can function as a tonic pronoun with the subject clitic *ce/c'* (often realized with *être* as *ça c'est…* 'that, it's…'), but also possible with subject clitic *ça*, as in *ça, ça depend des cours* ('that, it depends on courses'; cf. Thibault, 1983: 32).

Number agreement with *ce/c'/ça* is invariable in most contexts of colloquial French. *Ce* imposes default 3SG.MASC features on adjectives (e.g., *c'est beau* versus **c'est beaux* / **c'est belle*,

[66] Though *ça est* is not considered possible by Piron (2017), Morin (1982: 16) states that *ça est* is attested in Belgian French. *Ça est* can also be presumed to have been more widespread in the colloquial speech of older varieties of Hexagonal French, as seen in several dialogues from 17th and 18th century plays by Dancourt (e.g., *ça est bien surprenant* ['that is quite surprising'] in *Le galant jardinier*, 1704). The negated form *ça n'est pas* is much more commonly attested in contemporary colloquial French, due to the presence of the consonant [n] in *ne*-retention (cf. Morin, 1979: 23).

'it's beautiful'). In Modern French, verbal agreement with a plural anaphor as *ce sont* (13b), though recommended by normative grammarians, is considered a highly marked form in colloquial varieties and rarely attested (cf. Berrendonner & Béguelin, 2020); *c'est* has become the default form (13c). Furthermore, absent an anaphor, singular *c'est* is the default form, as in (14). As for *ça* functioning as a resumptive pronoun, this form always has singular verbal agreement, whether the referent is singular (15a) or plural (15b).

(13) a. *C'est la voiture qu'elle a achetée.* 'This is the car that she bought.'

b. *Ce sont les voitures qu'elle a achetées.* 'These are the cars that she bought.'
c. *C'est les voitures qu'elle a achetées.*

(14) *C'est difficile quand les gens me parlent en français.*
'It's difficult when people speak French to me.'

(15) a. *Paris ça a rien à voir avec la campagne.* 'Paris (it) is nothing like the countryside.'
b. *Les entreprises ça prend jamais de risque.* 'Companies (they/it) never take any risks.'

If in (15b) the subject were not doubled by *ça*, the verbal agreement would be determined by the NP *les entreprises*, giving *Les entreprises (ne) prennent jamais de risque.* Thus, when determining whether an NP may be in subject position rather than topic, it may be necessary to change the form of the verb to reflect agreement with the NP. I will finish this particular discussion by noting that further elaboration on the *ce/c'/ça/cela* distribution is beyond the scope of the current study (refer, e.g., to Piron (2017: 121) for a summary of *ce/ça* distribution and Ball (2000) for a pedagogical treatment of the distribution of *ce/ça/cela*; see also Kasper-Cushman (forthcoming) for a recent summary of these issues).

Let us return now to example 41 in Coveney (2003), reproduced as (16a). Here, the singular *est* is determined grammatically by the neuter pronoun *ce*. If *ce* were in fact an optional doubling clitic, it could be removed without a change in meaning. However, this results in the realignment of verbal number agreement with the NP for the sentence in (16b).

(16) a. *deux enfants c'est plus difficile* 'two children (it) is more difficult'
b. *deux enfants sont plus difficiles* 'two children are more difficult'

Here, it is the context that must determine whether *ce* constitutes a true doubling subject clitic. Coveney determines that the speaker did not, in fact, mean "Two children are more difficult." Rather, by looking at the immediately preceding clause, *avec un enfant ça va* ('with one child it's ok') one can reasonably assume (as Coveney does) that the speaker omitted repetition of the preposition *avec*, which would have made the topic status of *deux enfants* more explicit, as in (17). In this case, if the speaker had a contextual referent in mind for *ce*, or if one were required to assign a contextual referent, a subject NP such as *la situation* could be used as in (18), but obviously not *deux enfants*.

(17) *Avec deux enfants, c'est plus difficile*.
'With two children, it's more difficult.'

(18) *Avec deux enfants, la situation (c')est plus difficile*.
'With two children, the situation (it) is more difficult.

The following example from my corpus also illustrates how the choice of personal versus neuter pronoun can differentiate between SD and LD. On the topic of tennis players in (20), a native speaker (bilingual Fr) talks about how one player, Monfils, often shows emotion on the court; the speaker then switches to talking about another player, Federer.

(19) *[Monfils] externalise beaucoup. Federer, c'est pareil.*
'[Monfils] externalizes [his emotions] a lot. Federer, it's the same.'

As in Coveney's example, adding a preposition would have made the topic status of *Federer* more explicit: *Avec Federer, c'est pareil* ('With Federer, it's the same'); *la situation* as a contextual referent for *ce* would be appropriate for this example as well. This utterance can thus be considered an example of LD. However, inserting a co-referential personal pronoun instead of neuter *ce* would have been an example of SD:

(20) *[Monfils] il externalise beaucoup. Federer il est pareil.*
[Monfils] he externalizes [his emotions] a lot. Federer (he) is the same.

For the reasons advanced in Coveney (2003), and in Nagy et al. (2003) for L2 French learners, in the current study I will adopt this model treating all NPs followed by a co-referential subject clitic (whether personal or neuter, viz. *il/elle/ils/elles/ce/c'/ça*) as examples of SD, while excluding cases of LD such as topic shift where a subject clitic following an NP is clearly not co-referential (as in (19)).

5.1.4 Subject doubling and prescriptivism

From a diachronic perspective, there is evidence of subject clitics doubling the NP (whether the syntactic structure is considered LD or SD) in texts dating as far back as late Old French (13th c.).[67] Coveney (2003), citing work by Priestley (1955), states that disjunctive pronouns followed by a subject clitic (e.g., *moi, je…*) were present in Middle French and increased in use to the point that by the 16th century, they were as frequent as in contemporary French. Doubling of the subject NP, however, had only begun by the 16th century, and primarily with neuter *ce*; it wasn't until the 19th century that doubling with the personal pronouns *il/elle/ils/elles* became common.

Perhaps due to the fact that doubling of the subject NP was more marginal during the influential period of the 17th century grammarians, commentary on this structure as an "incorrect form" can be traced at least as far back as the grammarian Oudin, who wrote in 1632, "On ne met point de pronom personnel après un substantif, pour servir à un mesme sujet: par example on ne

[67] Vance (personal communication, July 21, 2020) provides the following example of SD involving *c'est*, appearing in 13th-century commentary on a French translation of the Bible (cf. Quereuil, 1988: 93):
"La lumiere qui fu faite le premier jor, ce est la foi qui est coumancement de toutes les vertuz."
La lumière qui fut faite le premier jour, c'est la foi qui est le commencement de toutes les vertues.
"The light which was made on the first day, this is the (Christian) faith which is the fount of all virtues" (English translation from Ayres-Bennett, 2005).

dit jamais, *Monsieur il a dit* mais *Monsieur a dit*"[68] (Oudin, 1632: 82). Similarly, Chifflet writes in 1659, "…ne dites pas *Mon père il est malade* au lieu de *Mon père est malade*"[69] (p. 59), and a guide by Le Touché instructing foreigners on how to speak French cautions against using doubling clitics, again for prescriptive reasons: "Ils disent, par exemple, *le Roi il est brave, la Reine elle est venue…Il, elle, ils, elles,* sont superflus dans ces endroits là et on ne les met avec un autre Nominatif"[70] (cited in Campion, 1984: 208). As Coveney (2003: 123) notes, based on observations by Blasco-Dulbecco (1999: 27-28), doubling with neuter *ce* seems to have been less stigmatized than with *il/elle/ils/elles*.

Campion (1984) speculates that this influence of the 17th century grammarians halted the process of grammaticalization of left-dislocation, leading to stylistically influenced variation in Modern French, where this structure is one of the more commonly cited characteristics of *français populaire* (see, e.g., Gadet, 1992: 70 (cited in Coveney, 2003): "On trouve presque toujours un pronom après le nom à la troisième personne"[71]). This prescriptivist influence may also be reflected through corpus analyses such as in Villeneuve and Auger (2013), whose bilingual French-Picard speakers demonstrated 25% SD in French versus 92% SD in Picard, where SD is stylistically more neutral. Note that, due to methodological differences in previous studies concerning what constitutes SD, it is difficult to make definitive conclusions that less-codified varieties of French demonstrate higher SD in direct comparison to varieties of Hexagonal French.[72]

---

[68] Campion's (1984: 207) translation: "One never places a personal pronoun after a noun, to serve as the same subject : for example one never says *The gentleman he has said* but *The gentleman has said*."

[69] "Do not say *My father he is sick* in place of *My father is sick*" (translation mine).

[70] "They say, for example, *the king he is good, the queen she has come…He, she, they* are superfluous in these positions and one does not place them with another noun" (translation mine).

[71] "One almost always finds a pronoun after a noun in the third person" (translation mine).

[72] For example, Auger and Villeneuve (2010) find 45% SD in Saguenay, Québec, lower than Sankoff (1982) for Montréal (55%) and Fonseca-Greber (2000) for Switzerland, but higher than Nadasdi (1995b) for Ontario (27%), Ashby (1980) for Tours (21%), and Coveney (2005) for Picardie (24%). As I discuss in section 5.2, differences in population demographics also likely account for much of this variation.

It seems, rather, that SD is widespread in contemporary informal spoken forms of French throughout the French-speaking world, even though the prestige of the written language may still influence the grammaticalization of certain features in contemporary varieties of French (see, e.g., Fonseca-Greber, 2000, who argues that the prestige of the written form seems to be the only factor halting the complete grammaticalization of SD in Swiss French). Dubuisson, Emirkanian, and Lemay (1983) observe this influence of standardization in a study of L1 French children, who use SD in both written and spoken form in the second year of primary school, but by the sixth year, SD in written form had decreased while in oral form it had increased as children acquired the sociolinguistic awareness of the non-doubled standard form (see section 5.2.1 for more discussion of SD in child L1 studies).

As Coveney (2005) notes, one of the preoccupations of early prescriptivist observers of language use was to avoid redundancy where possible, thus advocating for the avoidance of subject doubling. However, this rejection of redundancy is inconsistent, illustrated by prescriptivists' encouragement of maintenance of the standard form of complex interrogative inversion with NPs (e.g., *Mon père est-il malade ?* 'Is my father sick?'), which contains nearly identical doubling of the subject pronoun clitic. Thus, it seems that prescriptivists' preoccupations with redundancy only occurred when redundancy appeared in what were considered non-standard forms. Moreover, bipartite negation consisting of *ne*-retention in contemporary French likewise involves redundancy, at least on a semantic level (as discussed in Chapter 4, retention of *ne* has sociopragmatic functions, such as serving as a marker of a formal style, or marking emphasis on verbal negation in an informal style). The attitude on avoiding redundancy is nevertheless maintained in commentary by more recent grammarians, such as Moufflet (1935), who cautions against redundancy as a "contagious" aspect of spoken language; when addressing SD in

particular, Moufflet likewise laments, “Rien de tel que les mots superflus pour faire perdre le fil du discours et le sens du raisonnement”[73] (p. 202).

Coveney (2005) notes that prescriptivist attitudes seem to have softened somewhat even since the mid-20th century, but one can still find in popular contemporary media the notion that SD is to be avoided. There are examples of prescriptivists who seem to acknowledge the widespread existence of SD in colloquial French (*français populaire* or *français parlé* ‘spoken French’) and reserve the harshest criticism for its occurrence in planned discourse. Campese (2015) notes several uses of SD in speeches given by former presidents of France Nicolas Sarkozy and François Hollande, the latter having been the subject of multiple newspaper articles authored by prescriptivists criticizing his use of French in general (Combaz, 2014; Ferrand, 2015), including instances of SD (e.g., *La France, elle est…* ‘France, (she/it) is…’). One of the implications in these criticisms is that public figures such as Hollande are intentionally including features of colloquial French, such as SD, in order to tailor their language to an audience that would be receptive to such forms (i.e., showing the ability to use features characteristic of “popular” language common in the everyday speech of such an audience). Planned discourse aside, SD, like *ne*-deletion, does not seem to elicit the same notions of stigmatization as other informal discourse features such as interrogative structures, and Coveney (2005) notes that attitudes toward SD use in the classroom (L1 French) have likewise softened over the years, citing work from the 1970s (Dannequin, 1977) when teachers tried to suppress non-standard grammatical features such as SD, compared with more recent work (Guillon, 2002) indicating that teachers are more aware and tolerant of non-standard forms. Coveney (2005) summarizes the debate by quoting Blanche-Benveniste (1997:

[73] “Superfluous words risk losing the train of thought and the sense of reasoning” (translation mine).

98): “Despite the best efforts of prescriptivists, we can confidently assume that today *everyone* uses this structure at least occasionally.”

5.2 Studies on subject doubling in L1 French

The earliest corpus studies of subject doubling can be traced back to the 1970s. Though the number of published studies is lower compared to the number of studies on *ne*-retention, SD has been examined in similar French-speaking regions as for *ne*-retention (e.g., France, Belgium, Switzerland, and Canada). Table 5-1 lists these studies of SD for L1 French, in chronological order by year of survey.

Table 5-1. Studies on subject doubling in L1 French

Study	Year of survey	Research site	Speak-ers	Total tokens	SD% overall	SD range	Details
Ashby (1980)	1967-68	Paris	50	578	21.1	N/A	
Labelle (1976)	1971	Paris	20	N/A	79	N/A	5-year-olds (middle-class families); 4800 total utterances
		Montréal	20		89		
Campion (1984)	1971	Montréal	120	3384	36.4	N/A	Sankoff-Cedergren (1972) corpus
	N/A	Paris suburbs (Villejuif)	15	167	96.4	N/A	Adolescents from working-class families
Auger (1991)	1971	Montréal	2	N/A	N/A	25-30%	Sankoff-Cedergren corpus: upper class
			2	N/A	more than 70%	N/A	Sankoff-Cedergren corpus: lower middle class
Beauchemin & Martel (1977)	1971-74	Sherbrooke, QC	100	N/A	69	N/A	
Ashby (1988)	1976	Tours	25	383	73	N/A	
Sankoff (1982)	N/A	Marseille	2	77	84	82-89%	
		Montréal	4	189	48	33-60%	
Barnes (1985)	early 1980s	United States	3	389	79.2	N/A	L1 French speakers from France living in the U.S.
Nadasdi (1995a, b)	late 1970s	Ontario	117	2615	27	N/A	Mougeon corpus; adolescents
Coveney (2003, 2005)	1980	Picardy, France	30	1246	24.4	0-67%	
Auger & Villeneuve (2010)	1980-82	Saguenay area, QC	18	1479	45	18-79%	Paradis (1985) corpus
Beaulieu & Balcom (1998)	early 1990s	Northeastern New Brunswick	16	364	35.4	N/A	3SG + 3PL only
Fonseca-Greber (2000)	late 1990s	French-speaking Switzerland	14	1199	76.3	N/A	
Culbertson (2010)	2002-05	Lyon	5	1103	80.6	73-94%	Lyon corpus: Adult caregivers to children
			4	375	69.6	N/A	Lyon corpus: Children ages 1-4
Zahler (2014)	2005-06	Paris	17	1097	22	N/A	CFPP2000 corpus
Villeneuve & Auger (2013)	2006-07	Vimeu (Picardy)	8	529	31.8	11-65%	4 bilingual Picard-French; 4 monoling French
Palasis (2010)	2006-07	Nice area	20	938	53.2	N/A	CHILDES corpus; ages 2-3
Palasis (2015)	2007-08	Nice area	19	985	96.5	N/A	CHILDES corpus; ages 3-4

Where available, the *Total tokens* column and *SD% overall* column reflect the number of lexical NPs and the percentage of lexical NPs that co-occur with subject clitic anaphors. It is important to note, as Nagy et al. (2003) do, that direct comparisons of SD percentages across studies must be interpreted cautiously, since different methodological decisions were made in each case, and the "envelope of variation" is different for each one. As mentioned earlier, some studies include third-person strong pronoun subjects while others do not; in other cases, only subjects doubled by personal subject clitics are included, while doubling neuter clitics are excluded. Furthermore, some analyses characterize structures that fall under the definition of subject doubling (as outlined in section 5.1) in other terms such as left-dislocation (Ashby, 1988; Campion, 1984).

Possibly due to such differences in the scope of SD across studies, there is considerable variation in SD percentages. The remainder of this section elaborates briefly on particular sub-groupings of participants and research sites, followed by a more in-depth look at the factors influencing SD analyzed in these studies.

5.2.1 Subject doubling in children

Quantitative analyses of the L1 acquisition of subject doubling were conducted by Culbertson (2010) and Palasis (2010). As discussed in section 5.1.3, Culbertson analyzed children's speech in support of an argument for clitics as verbal agreement markers. As noted in Table 5-1, her data comes from the Lyon corpus (Demuth & Tremblay, 2008), which consists of approximately 106,000 utterances by four L1 French children and their caregivers. Culbertson notes the high level of SD use (81% overall) in child-directed speech by the four female caretakers, presumably due to some combination of the informal style of such speech and the extra information or facilitation of comprehension provided by a coreferential subject clitic. As for the children,

whose ages ranged between 11 months and 4 years 6 months, 70% of NPs (261/375) were doubled, with doubling of strong pronouns "at a high rate" (p. 118). Importantly, examined longitudinally, children's SD rates eventually matched or exceeded the SD rates used in their caretakers' child-directed speech. Another notable finding is the evidence of doubling with indefinite subjects (e.g., *un couteau il est là*, 'a knife (he) is here'), which Culbertson interprets as additional evidence that subject clitics are agreement markers in child grammar.

As part of a broader analysis of the acquisition of the subject, Palasis (2010: 293, 298) found somewhat less SD in children[74] aged between 2½ and 3 years than in Culbertson's study, with 72.4% SD (71/98) for strong pronouns *lui/elle/eux/elles* compared with 50.1% SD (428/840) for lexical NP subjects, though a later observation of children aged 3-4 in this longitudinal study revealed considerably higher SD rates at 96% for NPs (Palasis, 2015). Gotowski (2015), using Palasis' CHILDES corpus and focusing only on singular subject clitics directly preceding a strong pronoun or NP, compared children's speech with one adult interlocutor. She found that children had much higher SD (26.2%) than the adult (5.3%), from data including all singular persons (1st, 2nd, and 3rd person). However, her analysis was from the perspective of the percentage of subject clitics doubled with a strong pronoun or lexical NP (cf. Gadet (1997) in section 5.1), rather than the percentage of strong pronouns and lexical NP subjects doubled with a subject clitic pronoun. Since her focus was on subject clitics, she did not provide results of non-doubled NP subjects and strong pronouns, rendering direct comparisons with other SD studies in this section less applicable. Nevertheless, these results provide evidence that despite possibly considerable non-SD input from adults, children begin doubling with subject clitics early in their acquisition of French as their native language. These results also dovetail with observations of other sociolinguistic variables

[74] From Palasis' CHILDES corpus of 22 children in interaction with three adults at a preschool; two of the children were native speakers of languages other than French and were not included in the analysis.

(e.g., optional liaison, /l/-deletion, and /R/-deletion; cf. Nardy, Chevrot, and Barbu, 2014), where children's sociolinguistic usage at a young age (4-6 years old) converges with their peers toward non-standard variants, at an age when they already show awareness of standard sociolinguistic norms.

5.2.2 Non-Hexagonal French

This section briefly highlights some characteristics of subject doubling as it appears in non-Hexagonal French. SD is thought to occur in nearly all varieties of colloquial French, though what constitutes doubling has not been consistent in previous literature, as outlined in section 5.1.3. Outside of France, Canadian French is highly represented in studies on SD as compared to studies on *ne*-retention, likely due to the fact that SD is less categorical than *ne*-retention in most varieties of Canadian French. Note, however, that there is regional syntactic variation with differences in the subject pronoun paradigms of varieties of Canadian French (for example, the use of *je* for 1PL in some Acadian varieties, and the use of [al]/[a] for 3SG *elle* (cf. section 5.1.1) and *il* for 3PL with no overt plural or gender marking in Québec French, compared with 3SG [ɛl], 1PL *nous*, and 3PL *ils/elles* for Hexagonal French; see King & Nadasdi (1997: 269) for a brief overview). Swiss French also categorically uses *on* instead of *nous* for 1PL subjects, and Fonseca-Greber (2000) argues for the presence of *qui* as a 3SG doubling subject clitic.[75] Differences in the subject clitic and verbal inflection paradigms across regional varieties can thus create different phonological and morphological conditions in environments in which SD may be realized. Other regional morphosyntactic differences can influence SD patterns as well. For example, Hexagonal French is

[75] Fonseca-Greber (2000) motivates a treatment of an allomorph of *qui* as a doubling subject clitic in Swiss French. She considers the example *le patron qui me regarde* (lit. 'the boss who looks at me') to be functionally equivalent to Hexagonal French *le patron il me regarde* ('the boss (he) looks at me'), treating *qui* as a doubling clitic and not as a relative pronoun.

more likely to contain co-occurrences of SD and *ne*-retention (cf. Culbertson (2010: 95) with nine such tokens in her corpus of Lyonnais French), while such a sequence is expected to be exceedingly rare in Canadian French, given its near-categorical *ne*-deletion in informal spoken forms (cf. Auger & Villeneuve, 2010, with seven tokens of *ne*-retention and no co-occurrence of SD).

Stark (2013) conducted a study on the nature of (Swiss) French subject clitics in text messages, finding little use of SD, with only one token occurring with 57 lexical NPs. She thus posits that, if SD in European Colloquial French were already grammaticalized to an advanced degree as claimed by Culbertson (2010) and Fonseca-Greber (2000), such obligatory structures would appear more often than at a rate of 2% in text messages. Stark cautions that the written medium of text messages (with a focus on rapid communication, shortened orthographic forms and elimination of other elements unnecessary for comprehension) strongly inhibits SD, so a more accurate description may be that SD is a feature of "phonic" rather than "colloquial" French (p. 165), though she does not specify whether this characterization should be limited to Swiss French or can be applied to European Colloquial French (or to non-European varieties) more broadly.

Regional varieties have also been subject to discussions on whether SD in fact exists in certain varieties. Carroll (1982) argues that Québec French does not have SD, though subsequent work (Auger, 1994; Ossipov, 1990; Roberge, 1990) has challenged this conclusion. King and Nadasdi (1997) agree that SD occurs in Québec and Ontario French, but not in Newfoundland French, where instances of subject NP + clitic can be considered as either left-dislocation or as "separated" subjects (i.e., intervening pre-verbal material). Beaulieu and Balcom (1998) consider the Acadian variety of Northeastern New Brunswick to have subject doubling, reporting 35% SD in a study of 16 speakers.

Finally, the closely related language Picard (which is not always recognized as a distinct language by non-linguists) adds interesting data to the SD discussion, especially for bilingual French-Picard speakers. Auger (2003) and Villeneuve and Auger (2013) find, in general, high rates of SD in Picard and in more linguistic contexts compared with those in French. In Villeneuve and Auger (2013), bilingual French-Picard speakers averaged 92% SD when speaking in Picard (with one speaker demonstrating categorical SD of nominal subjects), compared with 25% SD in French; a monolingual French control group produced 42% SD. Furthermore, bilinguals produce SD much more frequently with quantified subjects in Picard (e.g., *Personne i n'vnoait*, 'Nobody (he) came,') compared with French (the equivalent expression *Personne ne venait* being marginal with a doubling clitic, viz., ?*Personne il (ne) venait*). Compared with the monolingual French speakers, the authors attribute the lower overall French SD rates among bilinguals to the fact that these bilinguals are actively involved in promoting Picard and have higher awareness of SD in both varieties, using SD to mark a clear contrast between French and Picard (rather than due solely to the prescriptivist influence of SD-less Standard French). Coveney (2005: 103) also identifies subject doubling as a possible "badge of Picard identity," so it is perhaps unsurprising that high SD rates obtain in Picard, and that such large contrasts are found in the SD rates of bilingual speakers across these two closely related languages.

5.2.3 Factors influencing subject doubling in L1 French

As one may expect in accordance with previous research on other sociolinguistic variables in French, quantitative analyses of SD show this structure to be conditioned by multiple linguistic and extralinguistic factors. In comparison with *ne*-retention, however, only a few studies on SD have been undertaken from a strictly variationist perspective: Nadasdi (1995b), Nagy et al. (2003), Auger and Villeneuve (2010), and Zahler (2014). Other studies, however, such as Barnes (1985),

Fonseca-Greber (2000), and Coveney (2003), also provide useful discussion on some of the factors influencing variation in SD. The following subsections will outline such factors identified as significant in influencing SD usage in such studies on L1 French. Though methodological decisions in the coding for each factor may differ across studies, it can be instructive to determine which factors are influential across these studies, allowing for more finely detailed analyses than comparisons based on overall SD percentages.

5.2.3.1 Linguistic factors

Linguistic factors influencing SD can be grouped into four broad categories. The first category is centered around properties of the (potential) doubling clitic pronoun, such as whether a personal or neuter pronoun is possible (section 5.2.3.1.1). The second category involves properties of the lexical subject, such as proper versus common nouns (section 5.2.3.1.2). The last two categories involve properties of the verbal argument of the subject: the third category concerns such differences as the type of verb (e.g., transitive verb versus copula; section 5.2.3.1.3), while the fourth category concerns the presence of preverbal material (that is, any intervening material between the lexical subject and the verb; this material may appear before or after the (potential) doubling clitic pronoun; section 5.2.3.1.4). The following subsections detail the properties of these linguistic factors.

5.2.3.1.1 Properties of the doubling clitic

Concerning the type of clitic that serves as the co-referential doubling pronoun, two earlier studies of L1 French are relevant. Both Ashby (1988) and Barnes (1985) find that, for lexical NPs that are doubled, the neuter clitic *ce/ça* is generally favored over personal clitics *il/ils/elle/elles*. Table 5-2 shows data available on this distinction in these two studies.

Table 5-2. Personal versus neuter pronouns in SD contexts

Study	% SD	Doubled NPs / Total NPs	Neuter subject clitics / Total doubled NPs		Personal subject clitics / Total doubled NPs	
Ashby (1988)	73.3%	281/383	170/281	60.5%	111/281	39.5%
Barnes (1985)	79.2%	308/389	227/308	73.7%	81/308	26.3%

Barnes (1985) takes this preference for SD with neuter subject clitics as evidence that doubling clitic *ce* has become grammaticalized in contexts where a lexical NP subject is followed by a 3SG/PL form of *être*. Likewise, Coveney (2003: 133) goes so far as to exclude such tokens from his data, suggesting that this structure of doubling (i.e., NP subject + *c'est*) does not generally have the same colloquial connotation as doubling with personal pronouns or with non-*être* verbs (where neuter *ça* would be used). Coveney nevertheless found 7% (23/329) non-doubling in 'NP + *(ce)* + *être*' contexts, indicating that doubling is not categorical, but may be in the process of grammaticalization. These considerations will be revisited later in light of the results from the current study.

Nagy et al. (2003) also consider the type of doubling clitic as a factor influencing variation in SD, dividing the clitic types into three contexts: 1) A possible doubling clitic must be personal (e.g., *Ma mère elle parle français*, 'My mother (she) speaks French'); 2) A possible doubling clitic must be neuter (e.g., *Paris c'est une ville sympa*, 'Paris (it) is a nice city'); or 3) A possible doubling clitic can be either personal or neuter—termed an "ambiguous" case where free variation may be possible (e.g., *Votre français il est très bon / Votre français c'est très bon*, 'Your French (he/it) is very good'). According to Nagy et al. (2003: 83), these distinctions in the coding were made according to French native speaker judgment.

Certain ambiguous cases (accepting either a personal or a neuter doubling clitic) appear to be in free variation. This is generally the case for singular inanimates. However, Donaldson (2008), citing Lambrecht (1981: 43), provides the examples in (21) and (22), noting that the choice of doubling clitic may indicate subtle contrasts in meaning.

(21) Les légumes, c'est où? 'The vegetables, it's where?'
(22) Les légumes, ils sont où? 'The vegetables, they are where?'

The first example can be interpreted to refer to vegetables as a general concept (e.g., a question asked by a shopper looking for the vegetable section of a store), while in the second example, the reference is to specific vegetables (e.g., a question asked by the same shopper who realizes she has misplaced the vegetables she just purchased). Donaldson also references Barnes (1985: 54) who indicates that neuter *ce* (in 'NP + *c'est*' structures) can denote a generic referent, though this interpretation is not categorical; *ce* "is open to virtually any sort of referent," whether individual entities or generic referents.

5.2.3.1.2 Properties of the lexical subject

Subject type has been demonstrated to be a robust predictor of SD frequencies across all previous variationist studies (Auger & Villeneuve, 2010; Nadasdi, 1995b; Nagy et al., 2003; Zahler, 2014). The types of subjects analyzed in previous studies on SD differ from the paradigm observed in Chapter 4 with *ne*-retention, so a brief explanation of the subject type paradigm concerning SD is warranted here. Lexical noun subjects divide into common nouns and proper nouns; common nouns can further be divided into definite and indefinite nouns, according to determiner type (e.g., *le* for definite and *un* for indefinite). Pronoun subjects include the strong pronouns (*lui/elle/eux/elles*) as well as indefinite pronouns such as *tout* ('everything/all'). Other subjects may consist of infinitives (e.g., *Être bilingue ça aide beaucoup*, 'Being bilingual (that)

helps a lot'). Finally, note that since subject clitics cannot themselves be duplicated, verbal arguments whose subject consists of a single subject clitic pronoun are not under consideration here.

Previous research on linguistic factors conditioning SD has found that other properties of the lexical subject favor SD, such as noun definiteness and specificity. The category of definiteness is rather straightforward; as Nadasdi (1995a: 5) specifies, [+definite] nouns include proper nouns, personal pronouns, and other nouns preceded by a definite article, while [-definite] nouns include nouns preceded by an indefinite article as well as traditional indefinite pronouns such as *quelqu'un, autre chose, plusieurs,* and *certains*. Specificity can be more difficult to identify, and there is necessarily some overlap with definiteness and specificity. Nadasdi again provides a succinct analysis; NPs that are [+specific] are those whose referent could be identified. Thus *mes parents* ('my parents') is [+specific], while *les gens* ('people'), though containing a definite article, is [+definite] but usually [-specific] (e.g., *en Ontario, les gens ils hésitent* 'in Ontario, people (they) hesitate'). Nagy et al. (2003) further divide [-specific] nouns into two categories, for nouns with indefinite articles as [-specific] and nouns such as *les gens* categorized as "generalizing," suspecting that the latter category would favor SD. Nadasdi (1995a, b) and Nagy et al. (2003) show that [+specific] and [+definite] subjects favor SD, though this distinction is not as robust in Auger and Villeneuve (2010), who found that while specific subjects favored SD, certain nonspecific subjects (generic subjects) favored SD the most. Nagy et al. (2003) also found that "generalizing" nouns favored SD at almost the same rate (45%) as [+specific] nouns (48%).

In the studies cited above, general trends that have been observed include strong pronouns favoring SD more than proper nouns, which favor SD more than common nouns. Indefinite NPs also tend to disfavor SD. In earlier studies, Roberge (1990) speculated that SD was not possible

with indefinite nouns (e.g., *un homme il est venu*, 'a man (he) came'), but Auger (1994) and Nadasdi (1995b) find corpus evidence to the contrary for Montréal and Ontario French, respectively. Indefinite *pronouns*, however, disfavor SD more strongly than indefinite common nouns. In his corpus, Coveney (2005) noted no SD with indefinite pronouns, such as *tout le monde* ('everyone'), *chacun* ('each one'), and *tout* ('everything/all'), or with the indefinite determiner *chaque* ('each'), such as *chaque personne* ('each person'). Though he notes that SD with *tout le monde* is attested in "very broad working-class French" (p. 103), the aforementioned contexts are nevertheless categorical in Coveney's corpus. Auger (1994: 97), however, specifies that doubling with indefinite pronouns appears frequently enough that they cannot be considered performance errors and therefore fall within the scope of variable structures. Nadasdi found that the SD frequency with indefinite quantifiers is more substantial (18%) in Ontario French than in Hexagonal French, though SD remains less frequent than with non-quantified subjects (27%). Furthermore, Nadasdi treats *tout le monde* separately from other quantified forms as the most disfavored quantifier due to its highly non-specific nature, reporting a 5% SD rate for this expression in Ontario French. Fonseca-Greber (2000: 347) likewise separates *tout le monde* (with categorical absence of SD in her corpus of Swiss French) from other forms of '*tout* + NP' in her corpus (with variable SD), proposing that *tout le monde* is sufficiently lexicalized so as to be perceived by her speakers as a bare quantifier in the same vein as *tout, rien,* and *beaucoup.*[76]

[76] Despite these results, it is possible that non-referential *tout le monde* with doubling subject clitic *il* remains salient in Hexagonal French due to its appearance in the titles of certain productions of popular culture in France, such as the Jean Yanne film *Tout le monde il est beau, tout le monde il est gentil* (1972) and other television series with similar names created by Yanne (*Tout le monde il est gentil*), as well as a song (*Tout le monde*) by the French singer Zazie, containing the lyrics "Tout le monde il est beau." Fonseca-Greber (2000: 347) identifies a similar popular media reference in the Swiss weekly news magazine *L'Hebdo*: "tout le monde il est beau, tout le monde il est créatif…" However, Fonseca-Greber's corpus results lead her to conclude that such media has not seemed to have any impact on speakers' lack of SD in this expression.

Quantified subjects overall have been shown to clearly disfavor SD. Villeneuve and Auger (2013) explain this phenomenon with reference to Rizzi's (1986) argument that quantified subjects cannot be dislocated. Thus, they argue that these subjects are among those least susceptible to SD, at least in Hexagonal French. Fonseca-Greber (2000) finds the same phenomenon in Swiss French, arguing that quantifiers should not be considered a qualitatively different class of NPs, contra Rizzi (1986) and Roberge (1990).

In this vein, Nadasdi (1995a: 8) finds a correlation between subject specificity and SD. Earlier research (e.g., Chesterman, 1991; Comrie, 1981; Quirk, Greenbaum, Leech, & Svartvik, 1972) established a specificity continuum for nouns and pronouns across languages. Nadasdi's corpus reveals that the more specific the NP, the more likely it is to be doubled. Based on SD rates, Nadasdi's continuum of specificity for subject type in French proceeds from most specific to least specific in the order seen below in Table 5-3.

Table 5-3. Subject doubling according to degree of specificity of the subject NP (reproduced from Nadasdi 1995a: 8)

Noun Type	**SD Tokens**	**% SD**
1st- & 2nd-person strong pronouns	N/A	100
3rd-person strong pronouns	145/195	74
Proper noun	51/118	43
Common noun	496/2187	23
Indefinite pronoun	14/115	12

First- and second-person strong pronouns are obligatorily doubled, while the remaining NP types show decreasing frequencies of variation according to decreasing specificity.

Fonseca-Greber (2000: 346) also proposes the existence of a continuum of referentiality within the broad category of indefinite pronouns, where quantifiers with higher degrees of referentiality (e.g., *chacun, quelqu'un*) are more likely to favor SD than those lower on the

continuum (*beaucoup, rien, personne, tout*). One could also extend this scale to include the 5% SD with *tout le monde* in Nadasdi's corpus, which was not included with his indefinite pronoun data. Nadasdi notes that, since specificity favors doubling, the lack of specificity in *tout le monde*, combined with a quantifier (even if lexicalized), may explain its quasi-obligatory non-doubling. Note, however, that fine-grained distinctions in referentiality with *tout le monde* are possible. Auger (personal communication, July 17, 2019) proposes distinguishing *Tout le monde sait ça* ('Everyone knows that') from *Tout le monde dans mon village connaît Jean* ('Everyone in my village knows Jean'), where the latter has a higher degree of referentiality than the former due to the PP modifier *dans mon village*, and thus may be more susceptible to SD. Though none of the studies analyzing doubling with *tout le monde* have made such distinctions, this proposal may account for some of the variation that has been noted.

Extending the scale yet further, one could also include subjects of negative polarity, which have the lowest SD frequencies. Note that while Picard commonly allows for negative subjects (e.g., *aucun, personne*) to undergo SD (cf. Villeneuve & Auger, 2013: 121), doubling with such quantifiers appears to be extremely marginal in all varieties of French, though still possible. Zribi-Hertz (2011: 242) implies that the example *personne il m'aime* ('no one (he) loves me') is attested in colloquial varieties, and Auger (1994: 98) provides evidence that native speakers judge doubling with *personne* as acceptable.

Subject animacy has also been shown to play an important role in SD usage. Previous studies have identified a basic animate/inanimate distinction while further dividing the "inanimate" category into material objects and immaterial concepts. In Nadasdi (1995b), animate subjects clearly favored SD (31%), while inanimate/material subjects (e.g., a table) and inanimate/immaterial subjects (e.g., an idea or place) disfavored SD (13% and 12%, respectively).

Auger and Villeneuve (2010) did not find such influence of animacy on SD rates, though they found effects in subject definiteness and specificity; specifically, non-definite, quantified subjects as well as non-specific subjects highly disfavor SD.

5.2.3.1.3 Properties of the verb

The type of verb following the doubled subject has been classified in different ways across previous studies. Initially, transitivity was the main distinction, but verb type has been further subdivided into transitives (accusatives), intransitives (unergatives and unaccusatives), passives, modals, and copulas. Nadasdi (1995b) motivates these distinctions by considering whether the status of the verbal subject as an *agent* of the verb versus a *patient* would influence SD. Nadasdi finds a clear trend in terms of SD and transitivity of the verb: 37% SD for accusative verbs, 19% for unaccusatives, and 7% for passives. The other variationist studies on SD obtain similar results across these verb types but reveal other more fine-grained distinctions. Auger & Villeneuve (2010) find that copulas favor SD and that modal verbs favor SD the least. Nagy et al. (2003) also found copulas favoring SD for native speakers (based on Nadasdi's data), as well as for L2 speakers, but this category was not selected as significant in a variationist analysis for the latter group. Zahler (2014) also found a significant effect for verb frequency: more common verbs such as non-modal *aller, avoir, faire*, and *pouvoir* favored SD, suggesting a similar process of grammaticalization as the one proposed by Barnes (1985) for *c'est*.

Clause type, briefly mentioned in Chapter 4, can be divided into matrix clauses, subordinate clauses (of which conditional clauses may be extracted as a subset), and relative clauses modifying noun phrases. Nadasdi (1995b: 113) finds another clear trend where matrix clause verbs favor SD the most (30%), followed by subordinates (20%) and finally relatives (17%). Auger and Villeneuve (2010) report a similar progression in Saguenay French, though they found

that conditional clauses beginning with *si* 'if' were slightly less favorable to SD (20%) than other subordinates (31%), but not as disfavorable as relatives (4%). In Parisian French, Zahler (2014) finds the same general trends: matrix clauses favor SD (27%) while all non-matrix clauses are unfavorable (12%); relatives and conditional clauses most strongly disfavor SD.

5.2.3.1.4 Pre-verbal / intervening material

The last category of linguistic factors concerns any material appearing between the lexical subject NP and the verb (also called pre-verbal material). In previous studies, this material has included object clitics, the negation clitic *ne*, postnominal adjectives and modifiers, adverbs, prepositional phrases, relative clauses modifying the lexical NP, appositions, parentheticals (e.g., *ben…* 'well…'), backchannel feedback (when the interlocutor interrupts the speaker between the utterance of the NP and the verb), and hesitations. Briefly, it is hypothesized that, when intervening material separates the lexical NP from its verb, there is a tendency to reestablish the link between the subject and the verb. This can be done in two ways: repeat the lexical NP or insert the subject clitic pronoun co-referential with the NP. Nadasdi (1995b: 116) argues that it is not so much the different syntactic structures that intervene between the NP and the verb but the simple fact that distance has been created between two elements from a discursive point of view; that is, relative clauses modifying an NP, despite not being syntactically separated from the head noun that they modify, tend to favor SD just as intervening adverbs do. On the other hand, Laurendeau, Néron, and Fournier (1982) note that hesitations, which concern linguistic performance, tend to favor SD more strongly than syntactic constraints, which concern linguistic competence.

Regardless of these differences, in nearly all cases, each type of intervening material has been shown to favor SD, supporting the argument that the distance created favors the reestablishing of the link between the subject and the verb. The exceptions concern negation particles and object

clitics. Regarding negation, it is in fact the *absence* of the negation marker *ne* in verbal negation that has been found to favor SD (Coveney, 2003; Nadasdi, 1995b; Nagy et al., 2003). This is unsurprising, as *ne*-deletion is a marker of informal speech and co-occurs with subject doubling; in other words, it is the co-occurrence of features marking an informal style that supersedes the distance factor here. While other studies such as Nagy et al. (2003) divide this negation factor into a binary distinction (the presence or absence of *ne*, regardless of whether verbal negation occurs), Zahler (2014) divides this factor group into three categories: no verbal negation (affirmative), *ne*-retention (negative with *ne*), and *ne*-deletion (negative without *ne*); in her study, this factor group was the most significant predictor of SD, with *ne*-deletion strongly favoring SD and *ne*-retention strongly disfavoring SD (affirmative sentences neither favor nor disfavor SD). As for other object clitics (that is, excluding the *ne* particle), only Coveney (2003) found a clear increase in SD when object clitics are present (48% SD, compared to his 24% global SD rate); Auger and Villeneuve (2010) and Zahler (2014) did not find a specific effect on SD rates for the presence of object clitics; and Nadasdi (1995b) found a marginal disfavoring effect for object clitics (22%, compared to his 27% global SD rate).

5.2.3.2 Extralinguistic factors

As previous studies have found (and as the current study found in Chapter 4) for *ne*-retention, extralinguistic factors are likely to influence SD as well, though they may not necessarily be the same factors for SD as those for *ne*-retention. Again, these factors may include demographic factors such as age, sex, social class or education, as well as other non-linguistic factors such as the portion of the conversation or the type of oral production task.

Age effects on subject doubling have been identified as significant in previous studies, though no clear trends appear to have emerged. Ashby (1980) collected his corpus in the 1960s,

and middle age groups (speakers aged 30-59) favored SD while the younger (ages 20-29) and older (ages 60-69) groups disfavored it. Ashby attributes the somewhat surprising disfavoring of SD by younger speakers to an effort by these speakers to be accepted into a favored social class, the Parisian establishment, which would favor conservatism on the part of these speakers. Auger and Villeneuve (2010), reporting on data collected in the 1980s, find the highest percentages of SD to be produced by their youngest group (ages 15-24). Conversely, Zahler's (2014) data on Parisians from the mid-2000s found that the oldest age group (age 56 and older) had significantly higher SD frequencies than the other two age groups (ages 31-55 and under 31), and while the gap in SD frequencies between the youngest group and the middle age group was much smaller, the youngest group produced the least SD. This leads Zahler to suggest that there may be a possible decline in SD in Parisian French, though this finding may also be explained by the conservatism of younger Parisian speakers, as speculated by Ashby (1980)—a phenomenon that may still be relevant some 40 years later.

Though Auger and Villeneuve (2010) found no significant overall effect of speaker sex, young women produced the most SD in their study, leading the authors to suggest ongoing linguistic change in this variable, spearheaded by young women. While Zahler's results support the findings that women favor subject doubling, her cross-tabulations provide evidence that age and sex act together regarding SD usage, with older women producing the highest SD rates and young men the lowest rates.

As with *ne*-retention, social class appears to play a role in subject doubling in L1 French. For Hexagonal French, Ashby (1980) finds much lower SD rates (21%) in well-educated Parisian French speakers (i.e., professionals and university students) compared with a mix of upper-middle and lower-middle-class Tourangeau French speakers; Coveney (2003) finds that working-class

speakers in Picardie produce significantly higher SD (42%) than middle-class speakers (19%).[77] A similar trend is found in Nadasdi's (1995b) analysis of the majority Francophone community of Hawkesbury, Ontario, where working-class speakers produce higher SD (54%) than lower-middle-class speakers (30%), who produce slightly higher SD than middle-class speakers (25%). In minority Francophone communities of Ontario, however, this factor is not significant; rather, speakers in minority Francophone communities as a whole produce lower SD rates than speakers in majority Francophone communities.[78]

Overall, it appears that there is considerable consensus on the factors that condition SD usage in L1 French, based on the variationist analyses that have been conducted across different regional varieties, even if overall SD percentages vary greatly across studies—again, likely due to differences in methodology and participant sampling. The next section discusses the comparatively few studies focusing on SD in L2 French, including factors influencing SD usage in learners, followed by a more in-depth account of learner exposure to SD in classroom language learning contexts and suggested pedagogical norms regarding its acquisition in learners.

5.3 Subject doubling in L2 French

Compared with *ne*-retention, subject doubling has received far less attention in research on L2 French. From an acquisitional perspective, a learner must identify the sociolinguistic contexts in which SD is possible and choose the appropriate co-referential subject clitic. One may presume

[77] Following conventions in Marceau (1977) and Coveney (1996), Coveney's 2003 study divides his middle-class speakers into an *intermédiaire* 'intermediate' group, consisting mostly of primary school teachers, and *supérieur* 'superior' group, consisting of middle school and high school teachers. The difference in SD rates between these two groups is small ('intermediate': 19.6%; 'superior': 17.5%), and Coveney states that, as far as sociolinguistic variables are concerned, few differences across these two groups would be expected anyway. I therefore include these two groups together as middle-class speakers.

[78] This difference led Nadasdi to include a subsequent extralinguistic factor, that of language restriction (cf. Mougeon & Beniak, 1991), which considers the frequency of use and range of registers used in the language. This factor was found to have a significant influence on SD (and likely on other sociolinguistic variables as well) for these types of communities: the greater the restriction, the less SD is produced.

that if SD is possible in the learner's L1, it may be an easier acquisitional task to incorporate this structure into the learner's L2 French, even if the sociostylistic functions do not necessarily match equally between the L1 and the L2.

Since the learners under consideration in the current study are L1 English speakers, a brief review of subject doubling in English is warranted. In both French and English, left-dislocation of NPs, including subject NPs, is possible, though in English the left-dislocation of subject NPs leading to the sequence 'NP + coreferential subject pronoun' tends to occur much less frequently than in French. Donaldson (2011a) specifies that in English, left-dislocation (which learners of French may interpret as superficially similar to SD) occurs primarily in unplanned, informal, interactional spoken discourse, or to introduce a referent inferable from previous discourse (Barnes, 1985; Geluykens, 1992; Gregory & Michaelis, 2001)—similar to LD and SD use in French, though in French this use occurs much more frequently in spontaneous as well as planned discourse. Nagy et al. (2003), analyzing SD in L2 French by L1 English speakers, observe that doubled subjects (termed "double marking") in English can, like French, be used for emphatic or contrastive purposes, but producing a doubled subject without emphatic or contrastive intent (that is, SD as it often appears in French) is much less common; in the English speech of their speakers, the authors observed a range of 0-5% of potential doubled subjects analogous to French SD (e.g., *My sister she's a music teacher*), compared to a range of 10-83% SD for the same speakers in French. The equivalent of French SD without emphatic or contrastive effect appears to be rare in most English dialects, though it does appear in certain dialects of Southern American English (Southard & Muller, 1998; Wolfram & Christian, 1976). Elsewhere, Tagliamonte and Jankowski (2019) find that analogous SD use in English may be due to influence from French SD patterns for

bilingual speakers, but in English, such apparent SD overwhelmingly occurs with proper names (and, to a lesser extent, animate subjects in general) compared with other types of subjects.

Returning to studies of SD in L2 French, some discussion of optional and required doubling with subject clitics is made in Towell, Hawkins, and Bazergui (1996), who analyzed the speech of advanced learners of French (L1 English, university students; minimum 6 months of study abroad in a French-speaking country) in a video retell task conducted before and after study abroad, in order to obtain measures of fluency. In the second iteration of the task, clitic doubling was observed with both lexical NPs and strong pronouns:

(23) a. *l'histoire ça commence avec* 'the story (it) begins with'
b. *lui il arrive* 'him (he) arrives' (Towell et al., 1996: 111)

Towell et al. include these doubling contexts among structures that, as a result of long-term residency in France, have become "proceduralized" (that is, lexicalized as "chunks") in the learner's oral production and available in short-term memory for use in rapid, unplanned speech.

Much of the research on structures that would include measures of SD usage in L2 French has focused instead on LD. Donaldson's dissertation (2008) and subsequent related work (2011a) concern the acquisition and use of LD by near-native French speakers. As mentioned in section 5.1.3, Donaldson's analysis does not distinguish LD from SD, and he notes that his study subsumes under LD some examples of what may have been considered SD in previous literature. He does provide LD figures as a percentage of all grammatical subjects (NPs and pronouns) eligible for topichood, though these rates are fairly low (range: 3-11%, including L1 speakers) if compared to previous studies focusing more specifically on SD in L1 speakers. While these details mean that his LD data are less comparable to the current study, one of his conclusions will be relevant: near-native speakers left-dislocate eligible subjects at a rate comparable to native speakers, suggesting

that near-natives are capable of nativelike mastery of the left periphery, and that they do not overuse such structures.

Other work characterized as LD in L2 French has mainly concerned earlier stages of acquisition. Earlier studies (Perdue, Deulofeu, & Trévise, 1992; Trévise, 1986) elicited spontaneous oral language through interviews. Trévise (1986) focuses on topicalization by L1 English and L1 Spanish learners of French. Though she provides few statistics, Trévise specifies that, over eight interviews, the L1 English speakers produced three occurrences of 'topic subject + anaphoric subject pronoun' (thus falling under the definition of SD developed in section 5.1.3), two of which are provided her discussion: *Reagan il a été très vague* ('Reagan (he) was very vague') and *Les petits garçons ils sont très plaisants* ('The little boys (they) are very nice'). Trévise notes that L1 English speakers use topicalization devices with a low degree of frequency except for lexicalized *c'est*, while L1 Spanish speakers use gendered co-referential subject clitic pronouns more frequently (including one example of doubling clitic *il* with *tout le monde*). Another study involving spontaneous oral language in interaction with native speakers is from Perdue, Deulofeu, and Trévise (1992), who conducted a longitudinal study of four learners of either L1 Moroccan Arabic or L1 Spanish, living in France and acquiring French largely outside the classroom. There is mention of LD being progressively acquired by all learners but no quantitative statistics are provided. Furthermore, the topic of conversation in this study was controlled in each instance since the elicitation methodology was a video retell task. More recent studies have illustrated that task type can strongly disfavor the production of LD (and by extension, SD). Through oral picture retells, Hendricks (2000) found only 6% LD with animate entities in L1 Chinese-L2 French speakers studying in France, and even less LD in native speakers (2%; compare with Donaldson's native speakers at 7.2%); Hendricks speculates that the task type led to a more formal style

disfavoring LDs. Likewise, Ferdinand (2002) had L1 Dutch-L2 French speakers and L1 French speakers perform oral picture narration tasks where cues for specific topic marking were triggered by events in the picture story (for example, requiring speakers to reintroduce or contrastively mark a topic), of which LD is one possibility. While the native French speakers did have higher LD rates (18% of all contexts; 5/28) than L2 speakers (10%; 10/101), and some of the examples provided by Ferdinand can indeed be considered subject doubling,[79] such a small corpus precludes any definitive conclusions. Ferdinand also cautions that the participants were in a formal experimental situation which, especially in the native speaker group, led to LDs being disfavored in favor of canonical SVO order. Sleemann (2004), replicating Ferdinand's elicitation methodology with oral picture narration, found nearly identical LD rates in L1 Dutch-L2 French speakers (19.4%; 18/93) and L1 French speakers (19.8%; 18/91), and an analysis of social factors revealed a correlation between frequency of exposure to native French in a natural situation and the use of LDs, suggesting that speech production in learners becomes more nativelike when they receive more native L2 input in a natural situation. Kerr (2002) also found that length of immersion experience in L1 English-L2 French university students corresponded to higher LD rates. However, despite a similar oral elicitation methodology (retells of a short film), native speakers in Kerr's study produced relatively little LD, rendering less instructive the results that may address whether learners are able to acquire nativelike competence in this domain. Given low LD rates by native speakers (that do not reflect the frequency of LD in otherwise unplanned, spoken French) in some of these studies, Donaldson (2011a) argues for an elicitation methodology that involves dyadic, turn-taking interaction.

[79] For example, an utterance by an L2 French speaker: *le première poisson il fait des trucs dans le mer* ('the first fish (he) does things in the water'). Note incorrect gender marking on *première*; cf. grammatical *le premier poisson.*

The studies described above are therefore of somewhat limited relevance to the question of SD as it concerns stylistic variation in L2 French speakers, and they highlight the need for spontaneous oral production tasks in both native and non-native speakers. At present, the study by Nagy et al. (2003) is the only full-scale analysis of subject doubling by L2 French speakers in spontaneous, dyadic conversation, and the only one to include a variationist analysis of L2 speakers, making it the most relevant comparison to the L2 data obtained in the current study. A basic description of the study follows in Table 5-4.

Table 5-4. Subject doubling in L2 French: Nagy et al. (2003)

Study	Year of survey	Research site	Speakers	Total tokens	%SD overall	SD range	Details
Nagy et al. (2003)	1993-94	Montréal	29	889	46	10-83%	Age range: 20-34

The L2 data in this study come from an earlier study (Sankoff et al., 1997) analyzing the use of discourse markers for L2 French speakers from immersion and non-immersion backgrounds. Sankoff et al. recruited L1 English speakers living in Montréal, and the speakers were recorded in an interview conducted by a native French speaker (one of the study authors). In Nagy et al., the speakers were ranked by an "acquisition score" which took into account the speakers' formal study of French and the type and amount of integration into the French-speaking community. These scores reveal much variation in the presumed proficiency of these speakers, from those who studied French as a formal subject in English-speaking elementary and high schools but have little contact with native speakers as adults, to those who were enrolled in French immersion schools and interact with native speakers on a daily basis as adults. The participants were self-selecting and agreed to be interviewed in French, for about one hour; thus, these speakers all presumably have at least conversational competency in French. The least fluent speakers

produced few discourse markers (2-3 markers in a 15-minute conversation sample compared with 40 for more fluent speakers) and had frequent pauses, whereas the most proficient speakers demonstrated nativelike use of discourse markers (which included French discourse markers *comme,* 'like,' *alors* 'so,' *bien* 'well,' *tu sais* 'you know,' *fait que* 'so,' *là* 'there,' *bon* 'good,' as well as English discourse markers used in French conversation, including *like, so, well, you know*). In Sankoff et al., the speakers were also assessed a grammar score determined by counting the number of gender errors in the first 20 non-ambiguous nouns uttered by each speaker and reporting the percentage of nouns that were correctly assigned for gender. Of the 17 speakers, the six highest grammar scores (at least 19 correct gender assignments out of 20 nouns) also corresponded to the six most frequent uses of French discourse markers. Based on the nativelike use of discourse markers in these speakers, as well as the correlation with conversational fluency and with gender accuracy, at least a subset of the participants in this study appear to be near-native speakers.

Furthermore, despite an apparently fixed role as interviewer versus interviewee, the topics of conversation included at least some elements not considered as "serious," such as family background, suggesting that informal discourse features may be favored (and expected) in this particular environment. In addition to the discourse markers mentioned above, all 17 L2 speakers produced at least 10% SD, although there were low NP token counts for several speakers (five speakers produced between three and seven total NP contexts for doubling). Higher use of the above discourse markers did not appear to entail high SD rates; only two of the six speakers with the highest use of discourse markers appear among the six highest SD rates, though only one of these six speakers produced SD at a rate (22%) well below the 46% overall SD rate.

Regarding linguistic factors, the L2 speakers' grammars appeared, broadly, to be sensitive to the same linguistic factors as L1 speakers' grammars (whose data was obtained from Nadasdi,

1995b). That is, all the significant factor groups conditioning SD in L2 speakers are significant in L1 speakers, and in most of these significant factor groups, the rankings of each factor also correspond between the L1 and L2 groups. Table 5-5 lists the significant linguistic factor groups for L2 speakers in Nagy et al. (2003), in descending order of significance.

Table 5-5. Significant factor groups for SD in Nagy et al. (2003: 90-91)

Factor group	L1 Ranking	L2 Ranking	Weight	%SD	Totals
Clitic type	Neuter > Personal	Neuter	.918	86	229/259
		Personal	.271	29	183/630
Subject type	Pronoun > Proper N > Common N	VP/PP	.904	95	18/19
		3SG.F pronoun	.886	79	23/29
		3SG.M pronoun	.817	68	40/59
		3PL.M pronoun	.653	48	16/33
		Proper noun	.620	63	66/104
		Common noun	.400	38	242/645
Other preverbal elements	Adverb > Hesitation	Hesitation	.786	61	30/49
		Parenthetical	.782	75	15/20
		Back-channel feedback	.759	64	14/22
		Adverb	.750	66	25/38
		PP	.631	55	31/56
		None	.434	41	290/704
Specificity	Specific > Nonspecific	Specific	.549	48	270/558
		Nonspecific	.427	24	17/70
		Generalizing	.416	45	118/261
Preverbal negator	Ø > *ne*	Ø	.517	46	403/869
		ne	.046	10	2/20
Clause type	Main > Subordinate > Relative	Main	.535	47	359/759
		Relative	.444	50	2/4
		Subordinate	.372	39	42/109
		si	.058	12	2/17
Definiteness	Definite > Quantified > Indefinite	Definite	.531	49	355/725
		Quantified	.362	20	24/119
		Indefinite	.308	33	9/27
		Not applicable	.479	94	17/18

From this table, we see that the type of clitic itself is the most significant predictor of SD. When the subject is doubled, and the context either requires or allows a doubled neuter pronoun (*ce/ça*), this clitic appears much more frequently (86%) than personal pronoun clitics appearing in contexts where doubling requires or otherwise allows personal pronouns (29%).

Properties of the subject appear in several places in this table; though they do not necessarily dominate over other elements of the clause (as the 2nd, 4th, and 7th most significant factors), on the whole, it is clear that the subject exerts a considerable influence on SD rates. Strong pronouns and proper nouns favor SD, while common nouns disfavor it. Subject specificity and definiteness favor SD, while nonspecific, quantified, and indefinite subjects disfavor SD.[80]

Concerning verbal properties, Nagy et al. find that the clause type generally influences SD rates in L2 speakers in the same way as in native speakers. Matrix clauses favor SD, while subordinates overall disfavor it, with *si* clauses most strongly disfavoring SD. Since very few relative clauses with potential SD contexts were produced, the 50% SD in these contexts (two subjects doubled out of four relative clauses) must be interpreted cautiously. On the other hand, Verb type was not selected as significant for L2 speakers. Recall that in L1 studies, copulas tended to favor SD, while intransitives (unergatives and unaccusatives) and passives disfavor SD. The same general trend holds true in Nagy et al.'s L2 speakers (57% for copulas, 28% for intransitives, 30% for passives), but it appears that other factors more strongly condition SD usage.

Intervening material between the subject NP and the verb also accounts for variation in SD use; as a whole, intervening elements favor SD (62%) compared with no intervening material (41%). The rankings between L1 and L2 speakers are not the same, though Nagy et al. explain that this is due to methodological differences with Nadasdi's (1995b) L1 data. Furthermore, while Nadasdi included *ne* with other preverbal clitics, Nagy et al. treated it as a separate factor group and found significant effects with *ne*, though not with the remaining preverbal clitics, contra Nadasdi.

[80] Nagy et al. (2003: 88-89) note that specificity and definiteness, moreover, favor doubling and agreement marking in other languages representing a variety of language typologies, such as Cairene Arabic, Zulu, Hungarian, colloquial English, and Porteño Spanish.

The remaining factor groups that are significant for L1 speakers but not L2 speakers are Animacy, Verb type, relative clause presence, and Preverbal clitic. As Nagy et al. note, the discrepancy may be due to slight differences between L1 and L2 grammars or methodological differences such as the inclusion/exclusion of certain factors within each factor group.

As for extralinguistic factors, Nagy et al. (2003) found that integration into the Francophone community generally correlated with higher rates of SD, though there was much individual variation. The speakers' recent environment (i.e., the language they use at work, and their current degree of integration into the community, versus a less recent environment such as their language exposure while in school) appears to have the greatest influence on SD: adult integration accounts for more than twice as much variation as any other social factor. These findings lead Nagy et al. to conclude that "Subject doubling is only really acquired by people who actually speak French with Francophones" (2003: 92), which goes a bit further than the findings in previous studies such as Towell et al. (1996) and those focusing on L2 acquisition of LD such as Sleemann (2004): it may not be sufficient for learners to simply be exposed to naturalistic input in order to acquire features of colloquial speech. Instead, engagement with native speakers in authentic communication (measures of which were not necessarily obtained, or at least not reported, in the previous studies) appears to be necessary for such acquisition.

5.3.1 Subject doubling in the language learning classroom

As discussed in section 5.1.4, teacher attitudes have become more tolerant of SD in L1 French speakers, but SD use in the L2 classroom has not received much attention. Coveney (2005) notes that some textbooks oriented specifically toward acquisition of colloquial French include examples of SD, such as in Rodrigues and Neather (2007). As may be expected, variable subject doubling is much less frequently encountered in more traditional classroom textbooks for L2

French learners, though obligatory doubling with strong pronouns does often appear. In her discussion of textbook depictions of subject clitics, Fonseca-Greber (2000) provides the example in (24) of a (constructed) dialogue from a first-year college French textbook (Ariew & Nerenz, 1989: 185), in which third-person and first-person strong pronouns *elle* and *moi* are doubled (the former being optional, the latter being required; underlined in Fonseca-Greber's text):

(24) *—Qu'est-ce que vous pensez de ces fleurs?*
—Ma camarade, <u>elle</u>, elle pense que ces fleurs sont jolies, mais <u>moi</u>, je préfère ces autres fleurs.
'—What do you think of these flowers ?'
'—As for my friend, *she* thinks these flowers are pretty, but *I* prefer those flowers.'

A more recent example of a first-year college textbook, *Chez nous* (Valdman, Pons, & Scullen, 2013), does not metalinguistically address the subject-doubling phenomenon in informal styles, though it introduces obligatory doubling of strong pronouns in (constructed) dialogues at the beginning of the first lesson and also advises learners to memorize certain fixed expressions in their entirety rather than trying to translate them literally, as in (25) below (English translations mine). After the subject pronouns are presented, learners are advised that *on* is often used instead of *nous* to mean 'we,' and another 'strong pronoun + subject clitic' form is presented in (26) (for (26)-(28), English translations are reproduced verbatim from the textbook and in italics). On the same page, the use of 3PL *ils/elles* for groups of males and females is explained through use of a 'lexical NP + subject clitic' anaphor (27); note that the textbook's English translations do not include a corresponding subject pronoun. In the same lesson, strong pronouns are introduced, and one of the functions described is for emphasizing subjects when providing a contrast, in which a 3rd-person doubled subject is given (28); note how the words are bolded in the original text.

(25) *Tu es de Paris ?*
'Are you from Paris?'
Non, moi, je suis de Montréal.
'No, (me,) I'm from Montreal.' (Valdman et al., 2013: 3)

(26) *Nous, on est de Lille.*
'We are from Lille.' (Valdman et al., 2013: 8)

(27) a. *Anne et Sophie, elles sont en forme.*
'Anne and Sophie are fine.'
b. *Jean-Luc et Rémi, ils sont stressés.*
'Jean-Luc and Rémi are stressed out.'
c. *Julie et Damien, ils sont occupés.*
'Julie and Damien are busy.' (Valdman et al., 2013: 8)

(28) ***Moi**, je suis de Lausanne, mais **lui**, il est de Saumur.*
*'**I**'m from Lausanne, but **he**'s from Saumur.'* (Valdman et al., 2013: 8)

Near the end of this first lesson, students are exposed to SD with inanimate lexical NPs in the form of imagined speech from speakers who have written a letter (English translation mine).

(29) *Mon adresse, c'est Case Postale 1602. Le code postal, c'est CH-1211…*
'My address (it) is Case Postale 1602. The zip code, (it) is CH-1211…'
(Valdman et al., 2013: 12)

Within this passage, other elements remain in their formal variants, e.g., *Il n'y a pas de code postal* ('There is no zip code') with *ne*-retention, and *Vous savez que Genève est en Suisse, n'est-ce pas?* ('You know that Geneva is in Switzerland, don't you?') with formal *vous*, avoidance of doubling with *Genève* and use of the tag *n'est-ce pas*, which is rarely used in informal spoken French.

While some of the examples in (25)-(29) could be considered as LD and topicalization rather than SD (especially given the isolated contexts and the constructed nature of the sentences), learners are nevertheless exposed, from the first lesson, to several examples of a subject clitic anaphor following a strong pronoun or lexical NP. In subsequent lessons, there are a few more examples of SD/LD that would be expected in informal speech, including the following examples in (30) used to illustrate that singular *il* and plural *ils* have identical pronunciation before a consonant:

(30) a. *Mon cousin, il joue du piano.*
'My cousin, he plays piano.'

b. *Mes frères, ils jouent au foot.*
'My brothers, they play soccer.' (Valdman et al., 2013: 55)

Here, the English translations are provided verbatim from the textbook. Learners may notice how *il/ils* are translated into English and note additionally that such structures are not necessarily preferred in (standard) English but that the wording reproduces the French structure, in contrast to (27). These examples all thus provide an indirect approach to modeling SD, though if learners notice these differences between English and French at this early stage, they would still need additional input, whether implicitly or explicitly introduced, to infer that stylistically felicitous SD tends to occur in informal rather than formal styles.

5.3.2 Subject doubling and pedagogical norms

To this end, Ossipov (2002) advocates for a pedagogical norm regarding the teaching of SD to classroom learners. Though previous work by Antes (1995) has encouraged the teaching of stylistically felicitous use of SD (as well as dislocated structures more generally, including right- and left-dislocated subjects and objects) to beginning students, Ossipov cautions against any unrealistic expectations of beginning learners using dislocation productively; indeed, my own anecdotal classroom observations concur that beginning to early-intermediate learners' productive use of SD is extremely rare. Ossipov advocates that teachers encourage students to use the most common form of dislocation, that is, subject doubling with an NP used to introduce a new topic or to make a contrast. Additionally, she suggests that SD be encouraged in specific types of subject NPs (2002: 176): 1) When the subject NP is a "heavy NP" or a complex NP; 2) When the subject NP is a coordinate structure; 3) When the 'NP + *c'est*' construction can be used. In the first context, it is presumed that intervening material (e.g., a relative clause) between the subject NP and the verb will render verbal agreement more difficult, not only due to the simple fact of having more

"material" separating the noun subject from its verb, as the (grammatical) utterance by native speaker CaF in (31) demonstrates, but also due to having competing nouns in the intervening material, as the (ungrammatical) utterances by SA learners in (32) demonstrate.

(31) *Les affaires qui sont devant la Cour de justice de l'Union européenne elles deviennent quoi ?* (CaF)
'The cases that are brought to the Court of Justice of the European Union (they) will become what?'

(32) a. *Le longueur des cours sont plus courts.* (2S)
'The length of the courses are shorter.'
b. *Le parti des îles pour le fromage sont trop petits.* (3S)
'The part of the islands for cheese are too small.'

From a sentence processing perspective, the addition of intervening material creates distance measured in terms of the amount of time needed to maintain the subject properties [±number] (and in some cases [±gender]) in working memory before marking those properties on the verb. The properties of any intervening competing nouns must be discarded in favor of the subject NP. This agreement with the closest noun is widespread in spoken forms of French and English and attested in written forms as well.

Ossipov's solution is to encourage learners to use the neuter pronoun *ce/ça* where possible. As discussed in section 5.1.3, since *ce/ça* grammatically has singular number regardless of referent, the question of [±number] is avoided, or simplified, by the use of *ce/ça* to double the subject NP, and this is possible (though not necessarily preferred) even when the subject NP is plural, as in (33).

(33) *Les voitures, ça bloque la route*.
'Cars, (they/it) block the road.'

Since *ce/ça* also lack gender properties as a neuter pronoun, the necessity for calculating [+feminine] properties in the VP is reduced. For example, learners may have encountered different gender agreement in the VP of subjects referring to the names of cities, as in (34a) and (34b).

(34) a. *Paris est beau.*
b. *Paris est belle.*
c. *Paris c'est beau.*
'Paris is beautiful.'

Gender properties of cities do not seem to be completely grammaticalized across native speakers but do exhibit certain trends. A recent commentary on this phenomenon in the popular press (in *Le Figaro*; Develey, 2017) indicates that the city of Paris being marked as feminine appears to be a more recent trend, while the *Académie française* maintains that masculine gender is required with the presence of a preceding adjective (e.g., *Le vieux Nice*), and the *Office québécois de la langue française* comments that gender of cities reflects orthographic constraints similar to the gender of countries (i.e., cities/countries ending in *-e or -es* tend to have feminine gender). In any case, doubling the subject with *ce* obviates the need for overt gender marking in the VP.

Some issues can be raised with this approach, however. Among the factors influencing subject doubling outlined in section 5.2.3, I noted that the property of subject animacy as [+animate] does not generally allow for doubling with neuter *ce/ça*, except in pejorative or other particularly defined contexts (cf. Pooley, 1996: 181).

(35) a. *Mes parents partent demain.*
b. *Mes parents ils partent demain.*
c. ??*Mes parents ça part demain.*
'My parents leave tomorrow.'

Learners, whether in the classroom or in the target language community, would need to recognize that neuter *ce/ça* has a much more limited scope in contexts of [+animate] lexical NPs such as in (35). Furthermore, learners also need to be aware of the allophonic and stylistic distribution *ce/ça* (and *cela*), as outlined in section 5.1.2. While learners would do well to incorporate into their interlanguage grammar such distinctions regardless of the question of subject doubling, properties of *ce/ça* nevertheless must be taken into consideration along with all other factors conditioning

subject doubling. Finally, gender agreement mismatches such as *la ville c'est belle* ('the city it is beautiful') and *c'est belle* without NP referent are attested in learner speech; even overt instruction of neuter *ce/ça* co-occurring with [+masculine] [+singular] may not suppress learner attention to agreement marking with a lexical [+feminine] NP.

The above concerns are not meant to dismiss the notion of teaching specific strategies of subject doubling for classroom learners. As Étienne and Sax (2009) outline in great detail, sociolinguistic variation is either entirely lacking, underrepresented, or incorrectly represented in many beginning and intermediate French textbooks. The idea of a pedagogical norm as reflecting actual language use while consistent with native speaker expectations (cf. Valdman, 1989) can be beneficial to learners wishing to achieve communicative competence. Nevertheless, it can be instructive to see whether more advanced learners living in the target language community use some form of Ossipov's strategy, whether or not these learners have received any metalinguistic input regarding subject doubling, from teachers or members of the French-speaking community.

The following section now reports on the results obtained in the current study with regard to subject doubling. As with *ne*-retention, it must be emphasized that learners wishing to integrate into the target language community should not necessarily "aim" for 100% subject doubling, as this frequency does not reflect native speaker tendencies more broadly. Furthermore, native speakers tend to expect more standard forms from L2 speakers (with the perception of the learner as an "outsider" and not necessarily "allowed" to use informal forms), leading to arguments for the learner to acquire a pedagogical norm such as that advocated by Valdman (1989). Nevertheless, it can be instructive to determine whether learners' performance approaches native speaker rates in informal contexts and whether learners are conditioned by the same linguistic (and/or extralinguistic/sociostylistic) factors as native speakers. Finally, if the interlocutor's language

background plays a role in *ne*-retention, as evidence from Chapter 4 suggests, it may also influence the use of SD in certain learner groups.

5.4 Results: subject doubling

All lexical NP and strong pronoun subjects from the corpus of spontaneous oral production obtained from the SA learners, Near-NSs, and their interlocutors were analyzed as possible contexts for SD. As with the *ne*-retention results, this section will discuss the descriptive statistics regarding SD frequencies for both groups of learners and their interlocutors, to be followed by a variationist analysis of SD in section 5.5. I begin with a description of the judgments made by native speaker raters concerning types of possible doubling clitics, to be following by a discussion of several methodological decisions regarding the inclusion or exclusion of certain contexts of SD.

5.4.1 Native speaker judgments on clitic types

Recall that in section 5.2.3.1.1 on properties of the doubling clitic, previous studies such as Nagy et al. (2003) used native speaker judgments in determining whether a personal clitic, neuter clitic, or either type could be used as a co-referential anaphor of the subject NP. As I am not a native speaker of French, I recruited two native speakers to serve as raters for determining the type of doubling clitic possible for each context. These judgments will thus allow for a comparison with previous studies concerning the personal/neuter doubling pronoun distinction such as in Ashby (1988), Barnes (1985), and Nagy et al. (2003). Moreover, these judgments will also inform my decisions on certain tokens that were ultimately included or excluded from subsequent quantitative analyses, including the variationist analysis in section 5.5. The raters and their judgments are described in this subsection.

I instructed each rater to read excerpts from the transcripts of spoken French from native and non-native speakers in the current study, where the relevant clauses were extracted for all contexts containing a lexical NP, an indefinite pronoun (including quantifiers), or a strong pronoun in subject position (n = 2963). For each context, any doubling clitics that were actually produced by participants were removed from the excerpts, in order to avoid possible biases toward the rater judging one type of clitic to be favorable over another. The raters thus judged whether a personal or neuter subject clitic could be used to double the NP or strong pronoun subject—regardless of whether SD of any kind was produced—using the abbreviations *pers* (for personal clitics), *imp* (for impersonal, i.e., neuter clitics), and *amb* (for ambiguous cases where either clitic type could be used), following the categories in Nagy et al. (2003), as well as *none* if the rater deemed a doubling clitic impossible. For the *amb* category, I also instructed the raters to indicate if one clitic type was preferable over the other in the given context. The raters were instructed to ignore grammatical errors (especially prevalent in SA learners) and to make any changes in gender/number agreement to allow for either clitic type in ambiguous cases. For example, in the clause *l'histoire est courte* 'the story is short' (with overt feminine gender agreement on the adjective *courte* due to the feminine noun *histoire*), if the rater accepted the neuter clitic *ce*, she could presume that the actual utterance contained SD and would have been uttered by the speaker as *l'histoire c'est court*, with the adjective *court* containing no overt gender marking.

Rater #1 was a native speaker of Hexagonal French from the Montpellier area, female, in her late 20s, with prior training in linguistics. She provided judgments for all 2963 tokens over a period of several days. To obtain a measure of inter-rater reliability, a second native speaker rater was subsequently recruited. Rater #2 was a native speaker of Hexagonal French from the Paris area, female, in her late 20s, with prior training in linguistics. She provided judgments for a subset

of the corpus for each clitic type (n = 319; at least 17 tokens from each group of participants were judged).

I presumed that potential SD contexts limited to personal clitics—those with [+animate] subjects (e.g., *ma mère elle parle français*)—would have high agreement across both raters, so I chose a smaller subset of these contexts (n = 49) to be judged by rater #2; these contexts also included quantifiers such as *tous les étudiants* ('all the students'). Three discrepancies were observed, all with quantified NPs where rater #1 judged personal clitics possible but rater #2 judged SD impossible. These three examples are produced in (36) below.

(36) a. *L'une d'elles elle est mutée maintenant à Paris.* (1L)
b. *...et beaucoup d'acteurs aussi l'ont...* (FrE)
c. *...et chacun il pioche dedans.* (ChF)

'One of them (she) is transferred now to Paris.'
'…and a lot of actors also have it.'
'…and each one (he) rummages around inside.'

Note that SD was produced in (a) and (c); however, rater #2 simply appears to be more conservative in accepting SD with indefinites such as *chacun*.

All other tokens judged by both raters (n = 270) had four possible ratings: personal doubling clitic, neuter clitic, either personal or neuter ("ambiguous") clitic, or no doubling clitic possible. For these tokens, the agreement between the two raters is less frequent (196/270, 73%), likely due to the fact that the raters had the choice of four options for these tokens. With one exception, the 74 disagreements were due to one rater accepting ambiguous contexts and the other rater limiting these contexts to either personal or neuter, or one rater allowing a possible doubling clitic and the other rater determining that SD was not possible. Such discrepancies may also be unsurprising given that the raters often did not have additional discursive context to determine generic versus individual referents (see section 5.2.3.1.1), so disagreements may have resulted

from one rater making a broader, generic interpretation. Rater #2 was also more conservative in disallowing SD in 12 contexts where rater #1 judged SD to be possible, while in only two contexts did rater #2 allow SD where rater #1 judged SD impossible (one context with *tout le monde* and the other with *certaines choses*).

The lone exception where one rater accepted only a personal doubling pronoun and the other rater only a neuter pronoun is in (37).

(37) *…mais LEA se trouve en fait à Roubaix.*
'…but LEA is actually located in Roubaix.'

The likely reason for this discrepancy is that rater #1 understood *LEA* as an acronym (where the speaker was referring to an academic department/discipline entitled *Langues Etrangères Appliquées*) and assigning *imp* (neuter) as the possible doubling clitic, while rater #2 assigned *pers* (personal), possibly interpreting *LEA* as someone's name (e.g., the female name *Léa*). Otherwise, in the remaining 269 contexts, rater #1 was more likely to accept ambiguous contexts (n = 115) whereas rater #2 was much less likely (only 66 contexts rated as ambiguous). These results again indicate that rater #2 was more conservative overall in terms of what type of doubling pronoun could appear.

Despite these differences, in the coding for the variationist analysis, it was necessary to choose one rater's judgment over the other when identifying the variants possible for each context. Ultimately, I retained all judgments made by rater #1, not only since this rater provided judgments for all tokens (and the analysis would thus have the internal consistency of the same native speaker), but also because the more liberal acceptance of ambiguous contexts indicates a *possible* choice of either doubling clitic type, rather than eliminating possible variation that is in fact accepted by another native speaker.

5.4.2 Inclusions and exclusions

I now turn to a discussion of certain kinds of tokens that were included or excluded from the quantitative results. First, note that only doubled subjects at the left edge of the clause were considered, as all previous studies discussed in sections 5.2 and 5.3 have done. Thus, examples such as *Elle est sympa, ma mère* ('She is nice, my mother'), where the lexical NP is right-dislocated, were not included. As discussed in the latter part of section 5.1.3, I also excluded tokens that were determined to be instances of left-dislocation, where it was not clear that the doubling subject clitic was a direct co-referent with the lexical NP (e.g., *Federer, c'est pareil*). This was somewhat problematic in the case of the SA learners, however. These learners often produced ungrammatical structures in contexts of left-dislocation and subject doubling, including verbal agreement errors, rendering the distinction between LD and SD more difficult. However, as discussed in section 5.1.3 on Nagy et al.'s (2003) and Coveney's (2005) motivations for LD/SD distinctions, and due to the fact that I compare native speaker data with learner data, I excluded from the analysis only those cases where left-dislocation (i.e., an NP followed by a non-coreferential clitic pronoun) could clearly be determined. While this methodological consideration renders comparisons of raw SD frequencies with previous studies such as Nadasdi (1995b) less straightforward, it allows for a more accurate comparison of native and non-native speaker data within my corpus, as well as for a more direct comparison with the lone quantitative study of SD in L2 French (Nagy et al., 2003), which did not distinguish LD from SD. Ultimately, in the case of the SA learners, five tokens were judged to be left-dislocation; their native and near-native interlocutor also produced seven tokens of left-dislocation (three from the near-native and four from the NS). In the Near-NS conversations, 44 tokens by Near-NSs, 2 tokens by near-native interlocutors, and 42 tokens by NSs were determined to be left-dislocation. These 100 exclusions

represent 3.3% of the possible 3063 lexical NPs and strong pronouns, leaving 2963 tokens of subject doubling contexts.

It is also important to note that, when treating learner data, one option is to simply assume a blank slate for learners and include all possible contexts for SD, which, when comparing overall SD rates with NSs, would likely include a large number of categorical items (either 100% or 0% SD, depending on the context). The other option is to conduct a more detailed analysis of categorical contexts and determine whether variation occurs in these contexts for learners. If learners pattern like NSs in categorical contexts, then we can more directly compare SD rates limited to variable contexts across participant groups. In the following sections, I provide justification for the cases where potentially categorical contexts (as defined in previous studies) were either included or excluded from the overall quantitative results.

5.4.2.1 Strong pronouns

I included in the quantitative analysis certain strong pronouns in subject position, as these pronouns are subject to variation in the presence or absence of doubling clitics. However, I restricted the scope to masculine third-person strong pronouns as in previous studies (e.g., Zahler, 2014). As discussed in section 5.1.1, in Hexagonal French, the feminine strong pronouns *elle/elles* are phonologically identical to their corresponding subject clitic counterparts. In some cases of putative subject doubling, pitch and emphasis can determine whether the sequence *elle elle* is in fact an occurrence of subject doubling rather than a repetition of subject clitic *elle*. Examined in this way, 16 tokens of 'strong pronoun *elle* + clitic *elle*' were identified in the corpus (7 by Near-NSs, 8 by NSs, and 1 by the NS interlocutor for SA learners). During the process of transcribing the oral data, one likely example of strong pronoun *elle* without a doubling clitic was identified. Uttered by native speaker KeF, this token (see (38)) appears in a context where the speaker, who

is a musician, explains that he is familiar with American country music artists, whereas his partner, though she is American, is not particularly interested in country music.

(38) *Je connais beaucoup d'artistes de country que elle ne connaît pas.*
'I know a lot of country artists that she does not know.'

Two intonational features serve to identify *elle* as a strong pronoun here: First, the speaker produces a hiatus (i.e., no elision) between the conjunction *que* and the pronoun *elle*; second, the speaker lengthens the pronunciation of *elle*. Moreover, the situational contrast that the speaker proposes between himself (*je*) and his girlfriend (*elle*) in this context further establishes that *elle* serves as a strong pronoun here.[81] However, in order to account for subject-doubling variation with all other instances of *elle* in the same way as for masculine *lui* versus *il*, it would be necessary to undertake a detailed acoustic analysis where each instance of *elle* in subject position is analyzed and a judgment regarding its status as strong pronoun or clitic is made. Such an analysis is beyond the scope of this study. Therefore, the previously identified 17 tokens of strong pronoun *elle*, as well as the token in (38), will be excluded from the quantitative results.

5.4.2.2 Indefinite pronouns

Coveney (2005) excluded all subjects with negative polarity (*aucun, personne, rien*) and all indefinite pronouns (e.g., *tout le monde, chacun*) as they were categorically not doubled in his corpus. In my corpus, none of the four negative subjects was doubled[82]; these tokens were not

[81] If the roles were reversed and the contrastive function were maintained, a strong pronoun is likely to be uttered in the relative clause: *Elle connaît beaucoup d'artistes de country que <u>moi</u> je ne connais pas*, 'She knows a lot of country artists that (me) I do not know.'

[82] These include two tokens of *personne* and two tokens of '*aucun* + NP':
Personne se parle (3L), 'No one speaks to anyone'
Personne pense que ça peut être…(KeF), 'No one thinks that can be…'
Aucun anglophone ne va pas [sic] pouvoir expliquer pourquoi (SaE), 'No Anglophone is going to be able to explain why'
Aucun cours se ressemble (ThE), 'No course is like any other'

included. As for indefinite pronouns, the results are more varied. *Chacun* occurred six times, one of which occurred with a doubling clitic (and uttered by a NS):

(39) *Chacun il pioche dedans.* (Ch)
'Each one (he) picks through it.'

Though this pronoun disfavors subject doubling, the doubling phenomenon is nevertheless attested in informal styles, such as in Fonseca-Greber's (2000) Swiss French corpus. These tokens of *chacun* are included in my results. Doubling with *quelqu'un* was more strongly rejected by my native speaker raters, but attested in my corpus and included in the results, with one doubling by a NS and one by a Near-NS out of 14 total tokens:

(40) a. *Quelqu'un qui sait quelque chose il vient en France pour les allocations.* (7P)
'Someone who knows something (he) comes to France for the benefits.'

b. *Tant que quelqu'un il veut venir à ta place.* (KeF)
'As long as someone (he) wants to come take your position.'

Quantifiers such as *tout* and *beaucoup* appear to follow a pattern. As bare pronouns in subject position, my native speaker raters judged that doubling with *tout* is not possible:

(41) *Tout dépend des commandes extérieures*. (KeF)
**Tout il dépend des commandes extérieures.*
'Everything (it) depends on outside orders.'

Of the 27 occurrences of *tout* as bare subject pronoun in my corpus, one was in fact doubled (with *ce*), again by native speaker Ch (in a context explaining that, for students studying languages, the local university's curriculum prepares them solely for careers in teaching rather than in industry):

(42) *Clairement tout, tout c'est pour l'enseignement.* (Ch)
'Clearly everything, everything (it) is for teaching.'

When *tout* is followed by a lexical noun, doubling is more frequently attested, and generally judged acceptable by native speakers:

(43) *Tous les acquis et les avantages qui existaient ils sont en train d'être sapés* (KeF)
'All the gains and advantages that used to exist (they) are being undermined.'

Fonseca-Greber (2000) found a similar pattern in her corpus: bare *tout* was never doubled, while '*tout* + NP' was variably doubled, with "heavier" NPs following *tout* more likely to trigger SD (similar to the above example for speaker KeF). In my corpus, subject doubling occurred in 23.6% (17/72) of '*tout* + NP' contexts across all participants. There is one important exception, however, with the expression *tout le monde*, as discussed in section 5.2.3.1.2. My corpus aligns with previous observations of *tout le monde* as a lexicalized expression disfavoring SD (cf. Nadasdi, 1995b: 119) rather than a '*tout* + NP' construction. Recall that for Ontario French, Nadasdi found 4.8% SD (6/126) with *tout le monde*, while all other quantified subjects (including *tout*, *plusieurs*, etc.) had 18.5% SD (55/298), and Fonseca-Greber (2000) found categorical non-doubling of *tout le monde* ($n = 12$), whereas '*tout* + NP' structures ($n = 10$) had variable SD (30%). In my corpus, the same trends obtained: for *tout le monde* ($n = 85$), there was a single instance of SD, while for all other occurrences of '*tout* + NP' ($n = 72$), there were 16 instances of SD (22.2%).

The single example of doubled *tout le monde* in my corpus warrants further explanation. The token was uttered by a NS in Lille (KeF, himself a native of a neighboring city, Douai, in the north of France), in the context of a discussion of the local variety used by the inhabitants of Valenciennes, another neighboring city: *Quand tu arrives à Valenciennes tout le monde il parle comme ça* ('When you arrive in Valenciennes everyone (he) speaks like that').[83] This example could be considered as a more referential use of *tout le monde* with higher specificity (as Auger posits), limiting the scope of *tout le monde* to the inhabitants of a single city, which could increase the likelihood of a doubling clitic. Another explanation is that, during this utterance, the speaker

[83] It is possible that this speaker used the plural subject clitic *ils*, as in *Tout le monde ils parlent comme ça* 'Everyone (they) speak like that', since *il parle* and *ils parlent* are phonetically identical. However, given that bare subject *tout le monde* has singular verbal inflection (e.g., *tout le monde est parti* versus **tout le monde sont partis*, 'everyone has left'), I provide the transcription of this speaker's example in its singular verbal inflection. Citations elsewhere (e.g., Zribi-Hertz, 2011: 236) follow this convention as well.

switched from his own accent, largely devoid of regional characteristics, to an accent containing stereotypical features of Picard, including the pronunciation of *ça* as /ʃa/ rather than standard French /sa/. As Villeneuve and Auger (2013) note, doubling with *tout le monde* (as well as with bare *tout*) is attested in Picard much more commonly than in French, particularly among younger speakers; furthermore, doubling with referential quantifiers in general is much more frequent. It is possible that this native speaker associates subject doubling with regional (or otherwise non-standard) varieties of French, whether due to awareness of stereotyped forms of the language or due to actual input from speakers of Picard (cf. Coveney's (2005: 103) mention of subject doubling as a possible "badge of Picard identity"). For these reasons, this token can be considered an exceptional case to the general observation that *tout le monde* strongly disfavors doubling. Indeed, in my corpus, there were few cases of *tout le monde* that could be considered as referential; there were no tokens of *tout le monde* with a PP modifier and just two tokens with an adverbial modifier (*ici*, 'here').[84] Thus, for nearly all tokens, the most likely interpretation is non-referential, with categorical non-doubling. Furthermore, doubling with *tout le monde* was judged as either unacceptable or extremely marginal by my native speaker raters, but possible with *tout* + NP. Thus, my quantitative analysis includes all occurrences of *tout* + NP but not *tout le monde* and bare *tout.*

As for *beaucoup*, a similar pattern occurs where doubling with bare *beaucoup* in subject position is not accepted by native speakers, while doubling with '*beaucoup* + NP' is considered more acceptable, though marginal.

(44) *Beaucoup vivent à Boca Raton.* (Fr)
**Beaucoup ils vivent à Boca Raton.*
'A lot (they) live in Boca Raton.'

[84] These two tokens are as follows: *Tout le monde ici doit parler français* 'Everyone here must speak French' (speaker 1S); *Tout le monde ici est un peu à l'inverse* 'Everyone here is a bit to the contrary' (speaker 3L).

(45) ?*Beaucoup de stéréotypes sur les Français, ça vient de Paris.* (AmE)
'A lot of stereotypes about the French (they) come from Paris.'

The two occurrences of bare *beaucoup* in subject position were not doubled and are not included in my quantitative results.[85] Of the 21 instances of '*beaucoup* + NP', two were doubled (both uttered by the near-native interlocutor with the SA learners); these tokens are included in the results.[86] Fonseca-Greber (2000) notes the same complementary distribution, albeit with only two tokens: one SD token with '*beaucoup* + NP' in subject position, one token of no SD with bare subject *beaucoup*.

5.4.2.3 Tensed complementizer clauses with *ce qui* / *ce que*

Examples of doubling with tensed complementizer clauses containing indefinite relative pronouns such as *ce qui* and *ce que* ('what') in subject position, as in (46), have been excluded from quantitative analyses in some previous studies (cf. Auger & Villeneuve, 2010: 74), citing categorical doubling.

(46) *C'que je veux faire c'est d'aller au cégep.* (Auger & Villeneuve, 2010: 74)
'What I want to do is go to vocational school.'

Auger and Villeneuve do not indicate how many of such tokens occur in their corpus. Zahler (2014: 363) references the above example from Auger and Villeneuve while mentioning that the CFPP2000 corpus only contains two tokens of such structures; surprisingly, neither occurred with a doubled subject, though she does not provide the context of these two examples in her discussion. Coveney (2003: 133) excludes all tokens of the sequence 'NP + *(ce)* + *être*' (of which (46) above

[85] In addition to the example in (44), the same speaker (bilingual Fr) produced the other bare *beaucoup* subject: *Beaucoup disent que le niveau était très mauvais* 'Many say that the level was very bad.'
[86] The other occurrence of '*beaucoup* + NP' with subject doubling: *Beaucoup d'Américains ils vont à Paris* 'A lot of Americans (they) go to Paris.' Another token by a Near-NS in Pau (4P) was considered an instance of left-dislocation and not included: *Beaucoup des élèves aux États-Unis, financièrement ils sont obligés de travailler* 'A lot of the students in the United States, financially they are required to work.'

is an example), regardless of whether *ce qui/ce que* introduce the NP, citing quasi-obligatory doubling in this sequence, following Barnes (1985). Though clausal *ce qui/ce que* may be followed by verbs other than *être* in the matrix clause, Coveney does not indicate that these tokens appear in his corpus.

Despite observations of quasi-obligatory doubling with clausal *ce qui/ce que* in informal discourse, my native speaker raters judged lack of doubling acceptable for clausal *ce qui/ce que* with matrix-clause verbs other than *être*, likely due to the fact that non-doubling of these structures is common in more formal styles. While both doubling and non-doubling are grammatically possible regardless of style differences, there appears to be near-categorical doubling in informal styles (especially when clausal *ce qui/ce que* is followed by the matrix-clause copula as in *c'est*) and variable doubling in formal styles.

In my corpus, there were 36 utterances of *ce qui* / *ce que* in subject position. All such occurrences were uttered during the Near-NS conversations; no such tokens were uttered by SA learners or their interlocutors. No other indefinite relative pronouns such as *ce dont* in a tensed complementizer clause were produced in the corpus. Of the 36 utterances, 33 consisted of a tensed complementizer clause followed by a matrix-clause copula; all of these involved *c'est* in the matrix clause, thus producing categorical doubling for clausal *ce qui / ce que* when followed by matrix-clause copulas. Of the remaining three, one clause was not doubled (in a context where a Near-NS describes acting in a Shakespearean play and forgetting his specific lines but inventing an appropriate response):

(47) *Ce que je suis arrivé à dire…avait l'air d'être Shakespeare.* (6P)
'What I managed to say…had the appearance of being Shakespeare.'

It is possible that the non-copula matrix-clause verb (*avait*) uttered here produced a more favorable context for non-doubling, though the fact that the speaker paused slightly would have, on the

contrary, favored doubling, as Nagy et al. (2003) found. The other Near-NS to produce a non-copula matrix-clause verb for these contexts also produced an intervening element, which likely influenced doubling in this context:

(48) *Ce qu'ils font dans la vie, leurs expériences, ça va influencer comment ils voient les choses*. (9P)
'What they do in life, their experiences, (it) is going to influence how they see things.'

Finally, the last token for these contexts includes the only non-copula matrix-clause verb produced by a native speaker:

(49) *Enfin ce qu'il disait ça ressemblait à ce que le monsieur avait dit.* (FrF)
'Anyway what he said (it) resembled what the man had said.'

With no intervening elements or pauses, this token is the clearest instance of doubling occurring with *ce qui / ce que* in non-copula matrix-clause contexts. Given the paucity of tokens, however, it is impossible, based on this corpus, to meaningfully quantify variability in NSs or Near-NSs when the matrix clause does not contain *être*.

For L2 French speakers, the possibility of transfer from L1 English (where doubling in equivalent structures is either ungrammatical or extremely marginal; cf. *What he said was correct* versus **What he said it was correct*) and the absence of doubling in more formal styles of French present a situation in which less doubling (and thus more variation) may be produced in tensed complementizer clauses followed by matrix-clause copulas. In my corpus, Near-NSs (and near-native interlocutors) nevertheless produced categorical doubling in such contexts, suggesting that they have identified the ubiquity of structures such as *ce qui…c'est*, at least in informal styles. As for SA learners, who are likely more susceptible to L1 transfer, no tokens of indefinite subject *ce qui / ce que* were uttered, so it is unfortunately impossible to determine the potential effect of L1 interference based on this corpus. In the quantitative analysis, I therefore excluded the 33

categorically doubled matrix-clause copula tokens, while retaining the three non-copula contexts (with neuter clitic *ça*) in the quantitative results.

5.4.2.4 Strong pronoun *ça*

In the corpus, 380 tokens of doubling of the strong pronoun *ça* were produced. Of these, 361 tokens were of the form *ça c'est* or *ça c'était*. As in previous studies (Nagy et al., 2003; Villeneuve & Auger, 2013: 118-19), these tokens were excluded from analysis due to obligatory doubling. All Near-NSs and interlocutors each produced at least one token of *ça c'est*, while five of the SA learners (and both of their interlocutors) also produced at least one token of this form. No non-doubled tokens of the ungrammatical copula form *ça est* were produced, suggesting that learners have acquired this structure as a lexicalized chunk. Note, however, that there were several tokens of *ça est* in composite forms (such as *ça est devenu* 'that became,' *ça est descendu* 'that came down,' *ça est arrivé* 'that happened'), which include four tokens by Near-NSs and four tokens by bilingual Ch.

There were also 19 tokens of the form *ça ça*, where the first *ça* can be analyzed as a strong pronoun and the second *ça* as an allomorph of neuter subject clitic *ce/ça*. (One of these tokens included the sequence *tout ça ça...*). Note that in some cases it was not clear whether the speaker actually produced strong pronoun *ça* followed by subject clitic *ça*, or whether the speaker simply repeated the subject clitic *ça* while searching for a subsequent lexical item. In any case, such tokens were not included in the quantitative results.

5.4.2.5 Ambiguous cases

The last category concerns excluded items where the presence of the doubled clitic itself could not be determined. In the majority of cases, this involved a potential doubling pronoun *ce*

immediately following a noun containing a /s/ phoneme in word-final position. For example, it was impossible to determine whether the speaker doubled the lexical subject *France* (word-final /s/) with *ce* (word-initial /s/) in *La France c'est un pays de râleurs* 'France it is a country of complainers' or did not (*La France est un pays de râleurs*, 'France is a country of complainers'). Another excluded case of phonological ambiguity is in the doubled utterance *Ma vie c'est fait ici* 'My life (it) is made here', which is phonetically identical to non-doubled but reflexive *Ma vie s'est fait ici* ('My life has been made here').[87] Finally, consider *Les gens ils vont pour le festival* ('The people (they) go for the festival') versus *Les gens y vont pour le festival* ('The people go there for the festival'). The phonologically identical *ils* and *y* render this utterance ambiguous, though the doubled interpretation is unlikely since non-doubled *Les gens vont pour le festival* lacks a PP complement of place (e.g., *Les gens vont à Douai pour le festival*, 'The people go to Douai for the festival'). However, this example was uttered by a non-native speaker (JeE) for whom the lack of PP complement of place may simply represent a non-targetlike deviation; moreover, in the same conversational turn several seconds later, JeE produced doubling with the same subject in *Les gens ils se lâchent* ('The people (they) relax'). Due to these considerations, the former token was excluded. Finally, in addition to ambiguity, overlapping speech or other noise interference sometimes made it impossible to determine whether a doubling clitic pronoun was present or absent. Seventeen such unclear tokens were excluded.

After accounting for the methodological decisions described above, the remaining items were analyzed for the occurrence or absence of a doubling subject clitic. The following sections

[87] Standard French requires gender agreement on the past participle *fait* (*Ma vie s'est faite*), which would phonetically disambiguate from *Ma vie c'est fait*; lack of gender agreement on past participles is common in spoken French. (cf. Audibert-Gibier, 1992; Gaucher, 2013).

detail the quantitative results of SD in the current corpus, beginning with SA learners and their interlocutors, followed by both groups of Near-NSs and their interlocutors.

5.4.3 SA learners: Overall results

In the SA learners' conversations, 557 subject NPs and strong pronouns were identified as potential subject doubling contexts. Table 5-6 provides the token totals and SD frequencies for all SA learners and interlocutors.

Table 5-6. Subject doubling in SA learner group

Speaker ID	SD tokens	Total NPs	% SD
1S	16	36	44.4
2S	4	42	9.5
3S	9	62	14.5
4S	20	35	57.1
5S	11	80	13.8
6S	0	45	0.0
7S	5	51	9.8
8S	33	65	50.8
SA learners overall	**98**	**416**	**23.6**
L2 near-native interlocutor	**49**	**82**	**59.8**
L1 native interlocutor	**44**	**59**	**74.6**

As one may expect based on previous studies, there is considerable range in SD rates in these SA learners, from 0% to 57%. If one excludes the highest outlier in Nagy et al., who produced only six tokens, the range in Nagy et al. is from 10% to 70%, skewing higher than in the current study but similar in terms of the difference between the highest and lowest frequencies (60% vs. 57%). In the current study, a clear distinction can also be made between two groups of speakers: those with "high" doubling (1S, 4S, and 8S) and those with "low" doubling (2S, 3S, 5S, 6S, 7S).

Proficiency as determined by *c*-test scores does not appear to correlate with subject doubling rates in SA learners, as Figure 5-3 shows.

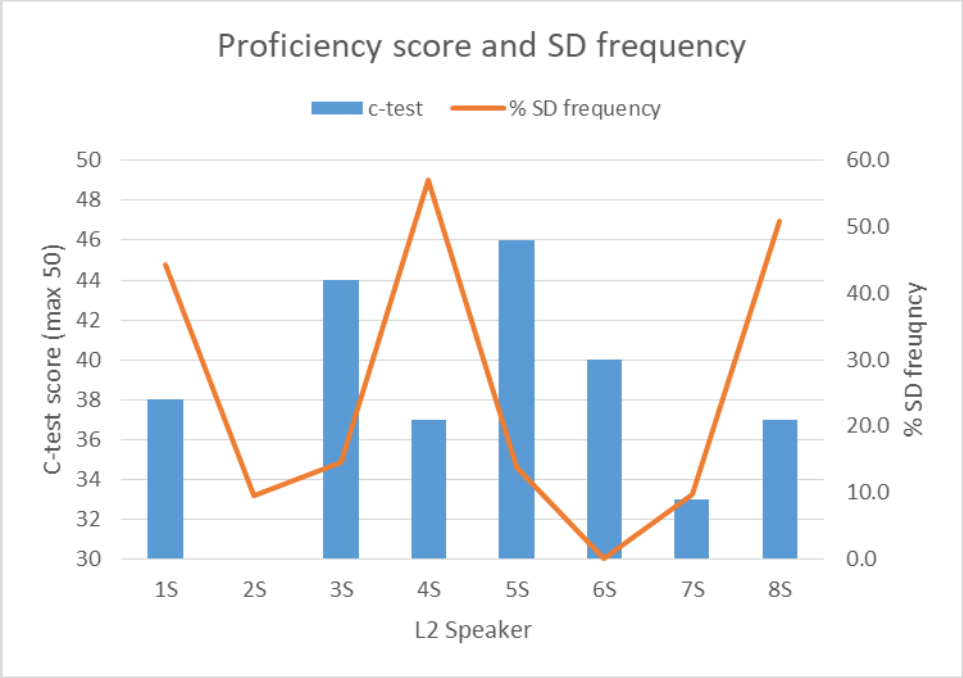

Figure 5-3. Proficiency score and SD frequency: SA learners

Note that the scale for the *c*-test scores (left side of graph) ranges from the lowest *c*-test score obtained in the SA learner group (30) to the highest possible score (50). From this figure, we see that the "high doublers" (1S, 4S, 8S) are in the mid-range for *c*-test scores (at 38, 37, and 37, respectively, essentially no different from the average *c*-test score of 38.1), while the three highest *c*-test scores actually correspond to lower SD rates.

On the other hand, language security, as measured by responses to the language background questionnaire, has a slightly better correlation with SD rates in these SA learners, as demonstrated in Figure 5-4.

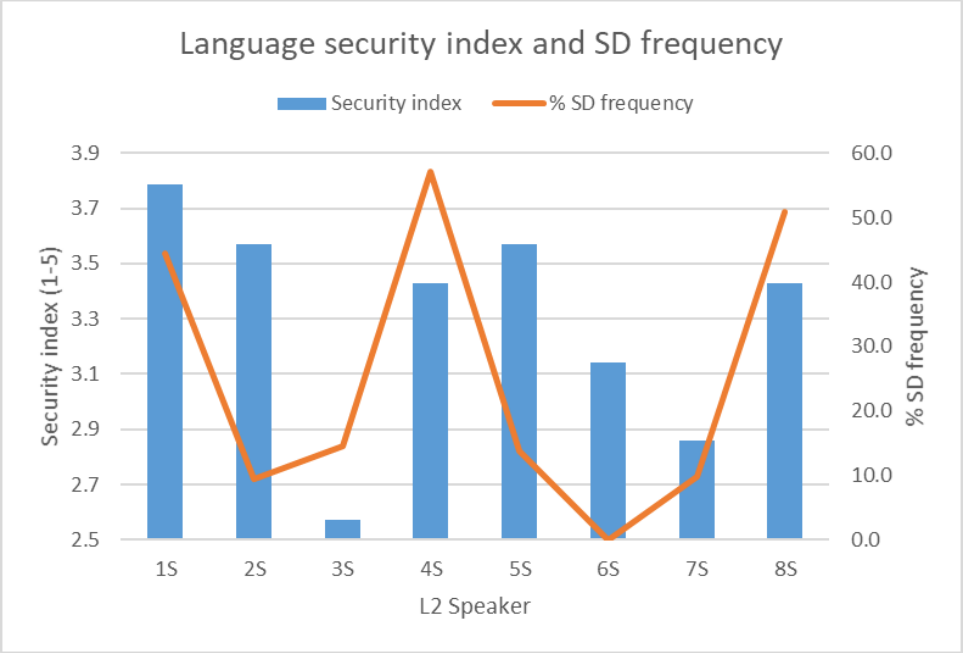

Figure 5-4. Language security index and SD frequency: SA learners

There are still two conspicuous outliers (2S and 5S), and one would expect a higher security index for 4S and a lower index for 6S. However, the three "high doublers" have language security indices higher than the average (3.3), while the remaining "low doublers" have lower than average security indices.

One possible factor accounting for the outliers with SD may concern the differences between self-rated proficiency measures and NS interaction. Recall that in Nagy et al. (2003), L2 French speakers with higher SD rates were the speakers who had more contact with NSs, especially as adults. As the language security index in the current study only includes self-rated proficiency measures, it may be instructive to examine other items from the language background questionnaire that focus specifically on social interaction with NSs. Four such questions were asked: one question on how often the speaker socializes with NSs in general, and three questions concerning how often the speaker interacts in French with NSs at home (likely to be frequent given these learners' homestay arrangement), with NS friends, and with other NS members of the

community. Though these measures are different from the integration scale devised in Nagy et al. (2003), and though SA learners in a short-term study-abroad program will not have benefited from an extended period of time for interacting with NSs in a variety of settings, it can be instructive to determine whether increased interaction with NSs correlates with higher SD rates for these speakers.

To do this comparison, the self-reported scores for these four questions were averaged[88] to produce a native speaker interaction score and plotted along with SD frequency. Figure 5-5 provides these results for SA learners.

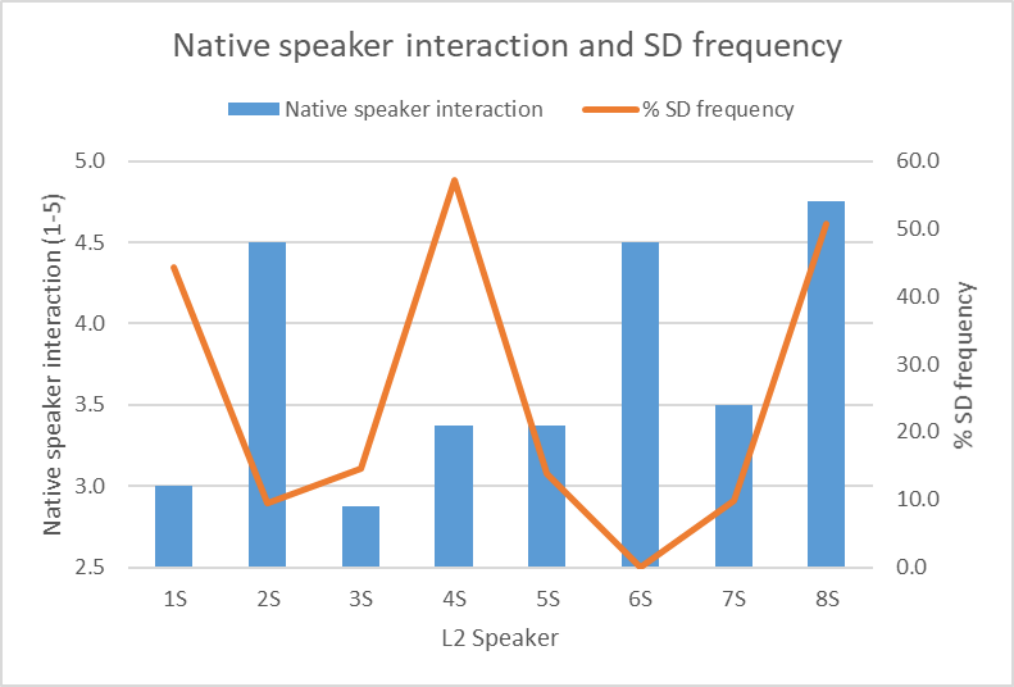

Figure 5-5. Native speaker interaction and SD frequency for SA learners

[88] Since the question reporting general socializing with NSs was administered on a scale of 1-4, this question was converted to a score out of 5 points. The other three questions were displayed on a scale from 1-5, with an additional category for providing details.

Immediately, it is apparent that this NS interaction score does not broadly correlate with SD frequency. Two speakers, 5S and 8S, indeed have more of a correlation than with the broader language security index, with 5S reporting lower NS interaction and 8S reporting higher interaction than the language security index, more closely reflecting their SD frequency. However, 1S and 4S, with high SD frequency for this SA learner group, have much lower NS interaction, while 6S, who produced no SD, reports higher NS interaction compared with overall language security.

There are multiple likely explanations for these discrepancies. First, the NS interaction scores are based on interactions that have been in place for a maximum of two months, given the time between the SA learners' arrival in France and the date of participation in the study. It is possible that benefits from NS interaction may not be immediately reflected in SD usage (cf. Castañeda & Zirger, 2011). Moreover, the homestay situation for these SA learners automatically increases their NS interaction score but does not reflect differences in the quality and quantity of interaction with their host families (cf. Hardison, 2014; Kinginger & Carnine, 2019). Thus, the results obtained in Nagy et al. (2003), finding that SD rates correlate with NS interaction, may be less applicable to lower-proficiency learners such as the SA learners in the current study, and this significant correlation may potentially be neutralized by the particular environment of these learners' short-term study-abroad program.

5.4.3.1 SA learners: Interlocutor results

I now turn to the SD results for the two interlocutors recruited to speak with the SA learner group (near-native AmE and native speaker SoF), who produced 142 subject NPs (third-person strong pronouns and lexical NPs) in all. The NS also produced one token of the strong pronoun *elle* with subject clitic *elle.* As discussed in section 5.1.1 and section 5.4.2.1, determining whether

all single utterances of subject *elle* or *elles* (that is, a context lacking SD) are strong pronouns or subject clitics would require an acoustic analysis beyond the scope of this study. Therefore, this token of the sequence 'strong pronoun *elle* + subject clitic *elle*' was excluded, leaving 141 NPs for subject doubling analysis. Table 5-7 shows that these interlocutors each produced substantially higher SD rates than the SA learners as a whole (23.6%), and these differences from the SA learner group were highly significant for both interlocutors (near-native: ($\chi^2(1) = 43.1$; $p < .0001$); NS: ($\chi^2(1) = 64.2$; $p < .0001$)).

Table 5-7. Subject doubling in native and near-native interlocutor (with SA learner group)

Speaker ID	**SD tokens**	**Total NPs**	**% SD**
Near-native interlocutor	49	82	59.8
NS interlocutor	44	59	74.6
Total	**93**	**141**	**66.0**

Both interlocutors' rates pattern similarly to overall frequencies obtained in other studies on SD in L1 French. Though the NS produced 14.8% more SD than the near-native, the difference between the SD rates in these two interlocutors was not statistically significant ($\chi^2(1) = 3.36$; $p = .067$).

5.4.4 SA learners: Subject doubling by interlocutor type

When dividing the SA learners' SD frequencies by interlocutor type, a clear pattern emerges, similar to that found in *ne*-retention. As Table 5-8 shows, SA learners have much higher SD frequencies in conversation with both the NS (SoF) and the near-native speaker (AmE), compared with SD use in conversation with another SA learner.

Table 5-8. SA learners' use of SD by interlocutor type

Interlocutor type	SA learners: SD / Total NPs	% SD
L1 French: NS	34/119	28.6
L2 French: near-native	53/172	30.8
L2 French: SA learner	11/125	8.8

As expected from these descriptive statistics, SA learners have significantly higher SD rates when speaking with the NS compared with other SA learners ($\chi^2(1) = 15.8$; $p < .0001$) and significantly higher rates when speaking with the Near-NS compared with other SA learners ($\chi^2(1) = 20.8$; $p < .0001$). There is no significant difference between SD rates in conversation with the NS versus the near-native ($\chi^2(1) = 0.169$; $p = .681$).

The degree to which these results mirror *ne*-retention results is striking, and this similarity across variables will be revisited in Chapter 6. On the whole, it again appears that SA learners treat the native and near-native interlocutor no differently with regard to certain sociolinguistic variables, despite the fact that these learners generally understood that the near-native was in fact a native speaker of American English. Furthermore, for both *ne*-retention and subject doubling, SA learners' linguistic usage is more classroom-like when interacting with other SA learners than with native and near-native speakers of French.

This distinction generally holds across individual variation in SD frequencies for these learners. Figure 5-6 plots the SD frequencies for each SA learner across each type of interlocutor.

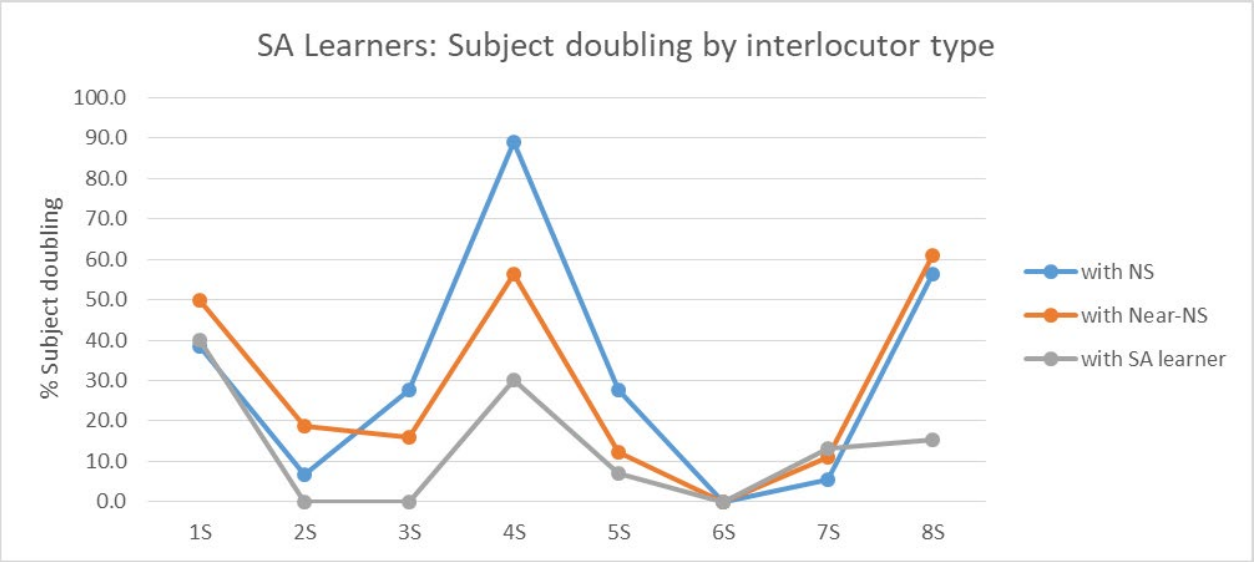

Figure 5-6. Subject doubling by interlocutor type for SA learners

Five of the eight speakers (2S, 3S, 4S, 5S, 8S) follow the same pattern as the group percentages indicate: SD rates are higher with both the Near-NS and NS than SD rates with another SA learner. Aside from speaker 6S who produced no SD, only one speaker (7S) had higher SD rates with SA learners than with the Near-NS or NS, though the actual differences in percentages are fairly marginal: 13.3% with other SA learners versus 11.1% with the Near-NS and 5.6% with the NS (8.3% when rates with the latter two interlocutors are combined). For the remaining speaker, 1S, his SD rate with another SA learner (40%) is lower than with the Near-NS (50%), but he is not included with the other five speakers since his SD rate with the NS (38.5%) was slightly lower than with the SA learner. However, this difference in percentage (1.5%) is extremely marginal, in addition to being conditioned by the low number of tokens in the SA learner-SA learner dyad (five; lower than any other total by interlocutor type). These considerations aside, as a group, these speakers exhibit remarkable similarity with regard to different interlocutor backgrounds, providing support for this factor as significantly impacting SD variation.

5.4.5 Near-NSs: Overall results

For the Near-NS groups and their interlocutors, 2,406 lexical NPs and strong pronouns were coded for the presence or absence of a doubled subject pronoun, with a global SD rate of 53.9% (1298/2406).[89] Table 5-9 gives the overall results concerning subject doubling for Near-NSs and interlocutors in Pau, followed by Table 5-10 for Near-NSs and interlocutors in Lille.

Table 5-9. Subject doubling in Near-NSs in Pau

Speaker ID: Pau	SD tokens	Total NPs	% SD
1P	17	81	18.2
2P	24	76	30.4
3P	24	41	62.5
4P	40	85	48.2
5P	21	88	28.4
6P	21	60	27.1
7P	31	68	39.2
8P	35	55	51.9
9P	40	89	43.7
10P	34	42	76.7
Overall Near-NSs	**287**	**685**	**41.9**
Bilingual interlocutors (English identity)	317	490	64.7
Bilingual interlocutors (French identity)	222	365	60.8
Overall bilingual interlocutors	**539**	**855**	**63.0**

Table 5-10. Subject doubling in Near-NSs in Lille

Speaker ID: Lille	SD tokens	Total NPs	% SD
1L	16	34	48.1
2L	35	52	69.8
3L	32	43	74.4
4L	26	46	56.5
5L	10	29	34.5
6L	13	44	29.5
7L	41	66	62.1
8L	14	25	56.0
9L	17	44	38.6
Overall Near-NSs	**204**	**383**	**53.3**
L2 near-native interlocutors	117	215	54.4
L1 native interlocutors	151	268	56.3
Overall interlocutors	**268**	**483**	**55.5**

[89] These figures do not include results from speakers who were recorded in conversation but excluded from the analysis: three speakers in Pau who did not meet Near-NS criteria, and one conversation between the near-native interlocutor (SaE) and the native speaker (CaF) recruited as interlocutors for the study in Lille.

Again, as with *ne*-retention, the variation in the range across Near-NSs is more evenly distributed, though the overall difference between the highest (76.7%) and lowest (18.2%) SD rates is quite similar to the range found in SA learners (58.5 across Near-NSs versus 57.1 across SA learners). Across both sites, SD frequency in Near-NSs is 46%, which is identical to the L2 speakers in Nagy et al. (2003). The order of conversations did not reveal a significant effect on SD for Near-NSs at either site, though in Pau higher SD in the first conversation (45.3%) than in the second (38.0%) was close to significant (Pau: $\chi^2(1) = 3.76$; $p = .053$; Lille: $\chi^2(1) = 0.740$; $p = .390$; overall: $\chi^2(1) = 0.806$; $p = .369$.

5.4.5.1 Near-NSs: Language background and subject doubling

When comparing the language security index and SD rate for Near-NSs at both sites, there is little obvious positive correlation between language security and SD rates. Figure 5-7 and Figure 5-8 show these comparisons for the Pau and Lille speakers, respectively.

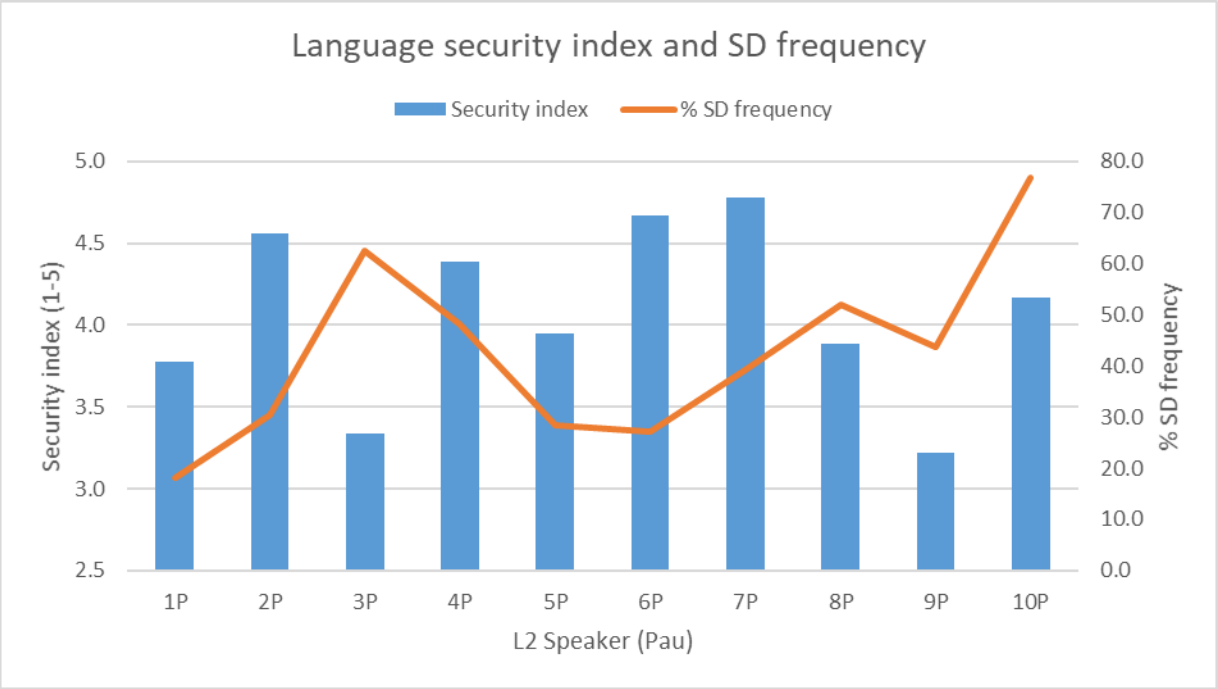

Figure 5-7. Language security index and SD frequency for Near-NSs in Pau

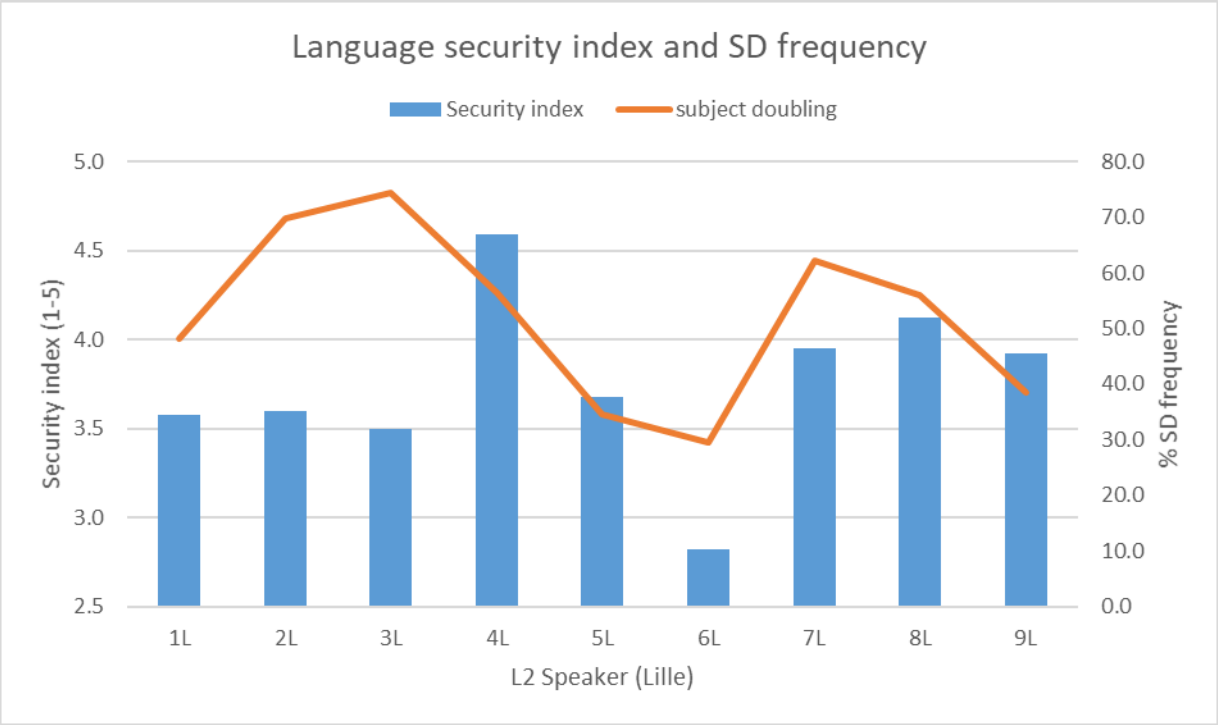

Figure 5-8. Language security index and SD frequency for Near-NSs in Lille

In Pau, the speakers with highest SD frequency (3P and 10P) are near the middle for language security, while the speakers with the highest language security (2P and 7P) are in the lower half of the group for SD frequency. In Lille, the correlation is slightly more robust; several speakers show some correlation between the two measures, and the speaker with the lowest language security also has the lowest SD frequency (6L), though this does not hold for the opposite end (i.e., the speaker with the highest SD frequency (3L) has one of the lowest language security indices; the speaker with the highest language security, 4L, is near the middle in terms of SD frequency). For Lille, when arranging the individual speakers by ascending order of language security, we see this general correlation more clearly applying to several speakers, as Figure 5-9 shows.

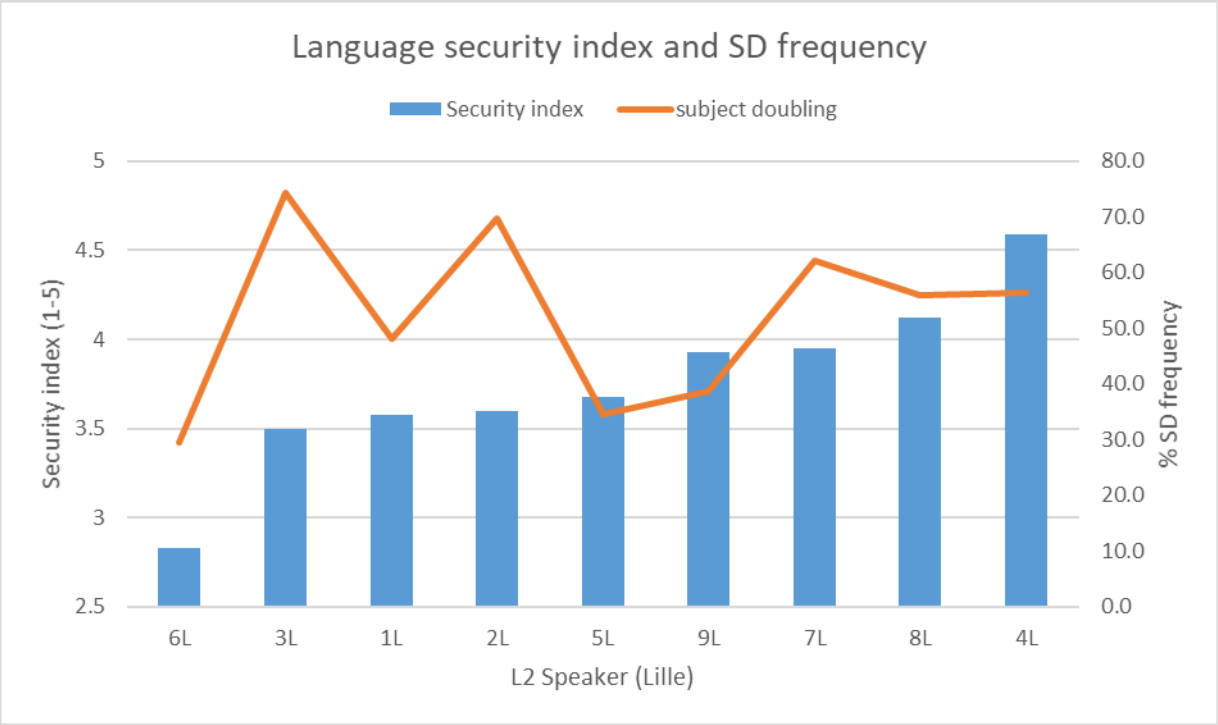

Figure 5-9. Near-NSs in Lille ordered by ascending language security

Speakers 2L and 3L appear to be clear outliers with regard to this general correlation. By interlocutor type, 3L had much higher SD (81.8%, 27/33) in conversation with her NS interlocutor than with her non-native interlocutor (50.0%, 5/10), though since 3L only produced 10 subject NPs in the latter conversation, the small token counts preclude definitive conclusions. The higher SD rate with the NS interlocutor may be due to a variety of factors, more of which will be examined in the variationist analysis (section 5.5). The differences in SD rates across interlocutors are not likely to be due to differences in social characteristics of 3L's interlocutors, since the NS and near-native interlocutors both had similar characteristics (i.e., younger, female) to 3L. Furthermore, 2L shares similarities in demographics with 3L and also had the same interlocutors, but with little difference in SD rates across conversations (64.3% with non-native versus 68.4% with NS). The possible interlocutor effect will be revisited in section 5.4.6. For speaker 2L, her relatively low security index is due partly to her responses that she rarely passes as a native speaker or attempts

to pass as a native speaker. For her, this behavior may not be relevant as far as sociolinguistic variation is concerned.

For SA learners, the analysis of a subset of the questionnaire involving interaction with native speakers did not find a correlation with SD rates (contra Nagy et al., 2003); however, given Near-NSs' presumed greater stability as members of the target-language community, it may be useful to compare Near-NSs' SD rates with this measure of integration with native speakers: how often the speaker socializes with NSs in general, as well as how often the speaker interacts in French with NSs at home (if applicable), with friends, and in the community. The same four questions were taken from the Near-NS questionnaires to determine a score from 1-5 measuring interaction with native speakers and were plotted along with SD rates. Figure 5-10 reports these results for Near-NSs in Lille.

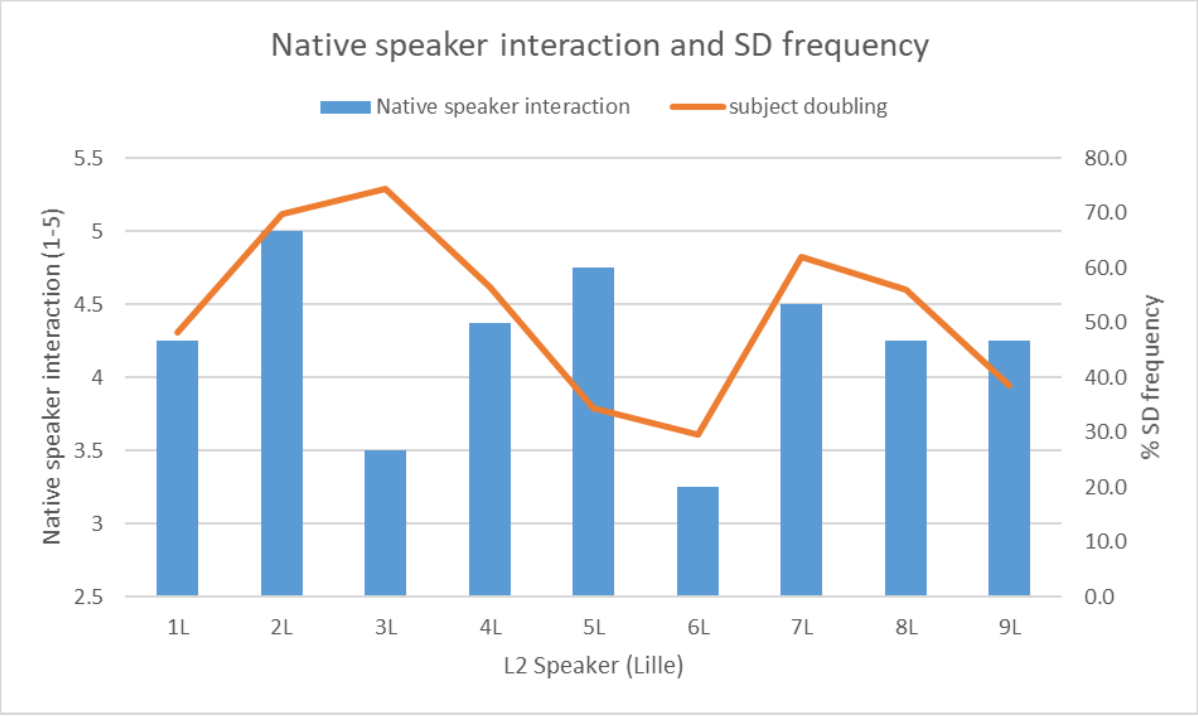

Figure 5-10. Native speaker interaction and SD frequency for Near-NSs in Lille

Here, we see a closer correlation with SD frequency than with the broader language security index, especially for 2L. Speaker 3L nevertheless remains an outlier. The primary reason for the lower NS interaction score for 3L is due to the fact that her partner is not a NS of French; otherwise, her interaction with NSs outside the home compares with other Near-NSs, though as an English teacher at the university level, she generally interacts with L1 French students in English, and she frequently interacts with L1 English colleagues in English. A possible explanation for her SD rates may come from a detail that 3L discussed during her dyadic conversations: she formerly had an L1 French roommate for one year, with whom she spoke in French. Had this detail been included in the questionnaire, her interaction score would have increased to 4.5, still lower than two other Near-NSs in Lille but producing a much higher correlation overall. Speaker 5L also presents as somewhat of an outlier. This speaker has had a lengthy residence in France (43 years), with a NS partner and daily interaction with NSs in the community, though she mentioned having fewer interactions in the community after retiring from her profession as an English teacher. Possible explanations for her lower SD rates may include the age gap with her interlocutors; as the oldest Near-NS in Lille, the age gap may have resulted in a higher degree of self-monitoring. However, this age gap did not seem to influence her *ne*-retention behavior (at 1.4%, second lowest of Near-NSs in Lille). Speaker 5L also produced the second fewest tokens of SD contexts (n = 29) in Lille; the low SD frequency may also be partly due to this low sampling rate.

As for Near-NSs in Pau (Figure 5-11), the same analysis of NS interaction and SD frequency produces little evidence of the general correlation that was found in Lille, though not to the same extreme as that observed in SA learners.

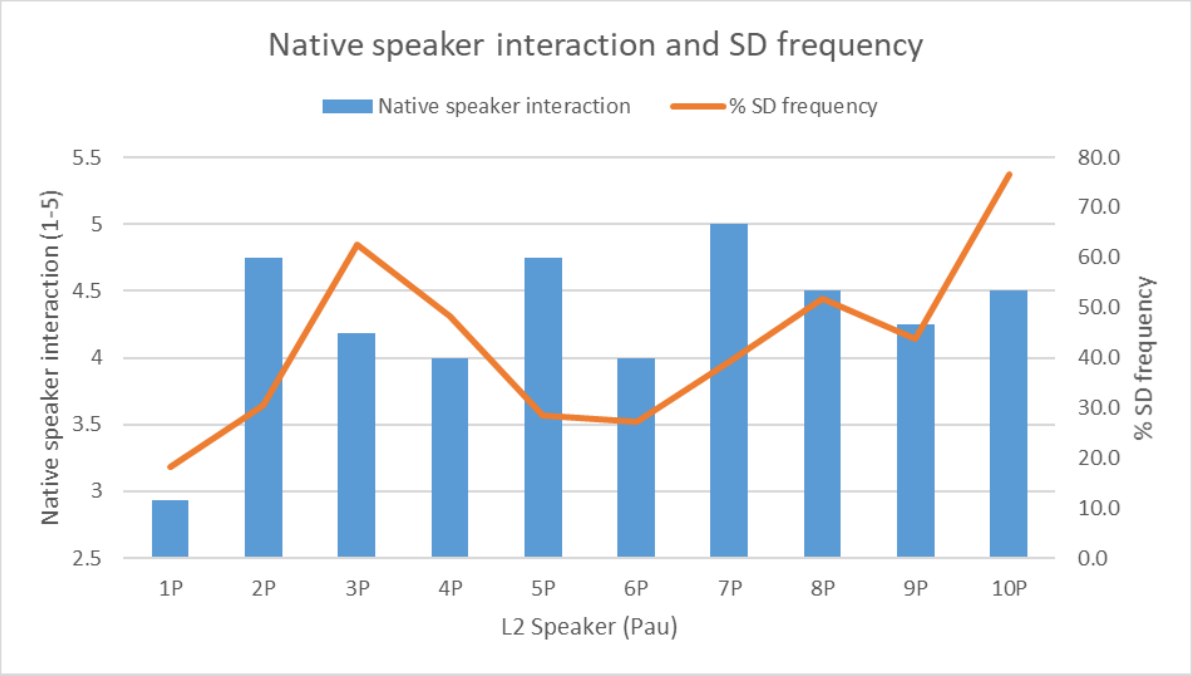

Figure 5-11. Native speaker interaction and SD frequency for Near-NSs in Pau.

One possible explanation of these findings for Near-NSs in Pau is that the older age of an L2 speaker may explain the lower (or lack of) correlation with NS interaction and SD frequency, since all speakers in Nagy et al. (2003) were between 20-34 years of age, and most Near-NSs in Lille were between these ages. Younger Near-NS speakers (all of whom began living in France as adults in the current study) may be more influenced by interaction with French NSs than older Near-NS speakers, who may be at more of a steady state (whether high or low) in their production of this sociolinguistic variable, even if their degrees of NS interaction differ.

These findings broadly support the conclusion obtained in Nagy et al. (2003), that increased contact with NSs correlates with higher SD rates, even though not all individual Near-NS speakers conform to this observation. Furthermore, this correlation appears to have more validity for speakers who are more "established" in the community and does not necessarily hold with SA learners who have had less time to establish contacts outside of their initial living arrangement.

Moreover, as noted in section 5.4.3, the questions pertaining to NS interaction did not provide for more fine-grained measures of contact with NSs, which may be particularly pertinent for SA learners on homestays.

5.4.5.2 Near-NSs: Interlocutor results

Turning to the SD results for interlocutors with the Near-NS groups, Table 5-11 provides the breakdown for the SD rates of each bilingual interlocutor in Pau, according to the identity they were asked to adopt for each conversation (note that *n* = number of conversations under each identity).

Table 5-11. Individual SD rates for bilingual interlocutors in Pau

Speaker ID	SD tokens	Total contexts	% SD
ChE (English identity; *n* = 4)	200	251	79.7
ChF (French identity; *n* = 5)	132	195	67.7
Ch (all)	**332**	**446**	**74.4**
FrE (English identity; *n* = 4)	79	190	41.6
FrF (French identity; *n* = 4)	77	128	60.2
Fr (all)	**156**	**318**	**49.1**
ThE (English identity; *n* = 2)	38	49	77.6
ThF (French identity; *n* = 1)	13	42	31.0
Th (all)	**51**	**91**	**56.0**
All bilinguals: English identity	317	490	64.7
All bilinguals: French identity	222	365	60.8
Overall bilingual interlocutors	**539**	**855**	**63.0**

As this table shows, across all three bilinguals, the SD rates according to identity are only marginally different, and this 3.9% difference is not statistically significant ($\chi^2(1) = 1.35$; $p = .246$). However, two of the bilinguals (Fr and Th), with fairly similar overall SD rates, have considerable differences according to identity. Again, as with *ne*-retention, sample sizes may explain these results. Recall that Th's one conversation under a French identity was with Near-NS 1P, with whom Th was most mismatched for gender and age. As with *ne*-retention, Th's SD rates were

more formal with this speaker. Table 5-12 shows that when including Th's conversations with participants who did not ultimately satisfy my criteria for Near-NS status (speakers 11P and 12P), his difference in SD average across identities is lower, showing that SD rates may not be as influenced by identity as much as the statistics in Table 5-11 would suggest.

Table 5-12. Individual SD rates for bilingual Th for each conversation.

Near-NS interlocutor and Th identity	SD tokens	Total contexts	% SD
1P (French)	13	42	31.0
3P (English)	28	36	77.8
4P (English)	10	13	76.9
11P (French)	14	15	93.3
12P (French)	11	12	91.7
English identity (with 3P, 4P)	38	49	77.6
French identity (with 1P, 11P, 12P)	38	69	55.1
Totals	76	118	64.4

With bilingual Ch, SD rates are comparatively high, and stable across all conversations save with 6P, as seen in Table 5-13. As I hypothesized for Th, Ch's lowest SD rate may be due to a mismatch in gender and age with 6P.

Table 5-13. Individual SD rates for bilingual Ch for each conversation.

Near-NS interlocutor and Ch identity	SD tokens	Total contexts	% SD
1P (English)	45	57	78.9
2P (French)	21	30	70.0
4P (French)	21	29	72.4
5P (English)	46	60	76.7
6P (French)	11	30	36.7
7P (English)	38	52	73.1
8P (English)	71	82	86.6
9P (French)	31	45	68.9
10P (French)	48	61	78.7
Totals	332	446	74.4

Bilingual Fr had the most unpredictable SD rates; there was considerable variation across conversations, as Table 5-14 shows.

Table 5-14. Individual SD rates for bilingual Fr for each conversation

Near-NS interlocutor and Fr identity	SD tokens	Total contexts	% SD
2P (English)	10	24	41.7
3P (French)	31	43	72.1
5P (French)	8	28	28.6
6P (English)	17	53	32.1
7P (French)	7	17	41.2
8P (French)	31	40	77.5
9P (English)	23	61	37.7
10P (English)	29	52	55.8
Totals	156	318	49.1

The only apparent explanation for this variation, based on non-linguistic factors, may lie in the ages of these Near-NS speakers: with the four Near-NSs under 50 years old, Fr's SD rate is 60.7%; with the four Near-NSs over 50 years old, his rate is 43.2%.

Overall, these bilingual speakers produce relatively high SD rates in comparison with their Near-NS interlocutors. These speakers also have relatively homogeneous social characteristics (similar ages and education levels), with the only major difference being gender. However, their Near-NS interlocutors vary widely in age and present themselves differently in terms of formality, despite the putatively informal environment (e.g., some Near-NSs used *vous* with the bilinguals and vice-versa). It is possible that the bilinguals were sensitive (whether consciously or subconsciously) to these differences with respect to their use of SD, adopting a more formal tone with speakers with whom they considered a more standard form of the language to be appropriate.

Turning now to the interlocutors for all conversations in Lille, Table 5-15 provides the descriptive statistics regarding their use of SD with Near-NSs.

Table 5-15. Subject doubling in native and near-native interlocutors (with Near-NSs in Lille)

Speaker ID	SD tokens	Total NPs	% SD
SaE	75	158	47.5
JeE	42	57	73.7
L2 near-native interlocutors: Lille	**117**	**215**	**54.4**
CaF	101	166	60.8
KeF	50	102	49.0
L1 native interlocutors	**151**	**268**	**56.3**
Total: all interlocutors (Lille)	**268**	**483**	**55.5**

The difference between the SD rates with the near-native interlocutors (SaE and JeE) is significant ($\chi^2(1) = 11.6$; $p = .001$); however, their combined SD rate is quite similar to the NS interlocutor rates (no significant difference; ($\chi^2(1) = 0.179$; $p = .672$)). The 11.8% difference between the two NSs (CaF and KeF) approaches significance ($\chi^2(1) = 3.59$; $p = .058$). In general, then, the interlocutors recruited in Lille present their Near-NSs with more variation in SD than in *ne*-retention, despite the fact that the overall SD rate across both types of speakers is quite similar.

5.4.6 Near-NSs: Subject doubling by interlocutor type

I now turn to the performance of Near-NSs as it relates to SD rates according to interlocutor type. Table 5-16 provides the SD results for Near-NSs in the current study, separated by data collection site and divided by interlocutor type (see third column). Results obtained by Nagy et al. (2003) are reported for the purpose of comparison.

Table 5-16. Subject doubling rates in Near-NSs divided by interlocutor L1 status

	Overall results for L2 French			**L1 and L2 results divided by L1 status**			**L2 speakers: results across interlocutor type**		
	SD	Total NPs	SD%	French status	SD / total NPs	SD%	Interlocutor type	SD / total NPs	SD%
Nagy et al. (2003)	405	889	45.6	L1	N/A	N/A	L1 French	405/889	45.6
				L2	405/889	45.6			
Near-NS: Pau	287	685	41.9	L1	539/855	63.0	L1 French identity	145/345	42.0
				L2	287/685	41.9	L1 English identity	142/340	41.8
Near-NS: Lille	204	383	53.3	L1	151/268	56.3	L1 French	136/234	58.1
				L2[90]	321/598	53.7	L2 French	68/149	45.6

Near-NSs in Pau essentially produce identical SD rates in conversation with bilinguals adopting an L1 French identity compared with an L1 English identity (a highly non-significant difference: ($\chi^2(1) = 0.005$; $p = .944$). However, in Lille, the Near-NSs produced significantly higher SD rates in conversation with NSs compared with near-native interlocutors ($\chi^2(1) = 5.70$; $p = .017$). This suggests that these speakers in Lille, despite being exposed to slightly more SD in their L1 English interlocutors (56.3% versus 53.7% in L1 French interlocutors), are less nativelike with regard to SD in their conversations with other near-natives. Both of these near-native interlocutors, as discussed in Chapter 3, meet the criteria as near-natives; yet their non-targetlike phonology and (morpho)syntax clearly identify them as non-native speakers. For Near-NSs in Lille, *ne*-retention did not appear to be influenced by these differences in fluency between near-native and native interlocutors, whereas SD does appear to be influenced, with this influence trending in the same direction for Near-NSs as for SA learners (i.e., higher SD with native speakers than with non-natives). This is the first piece of evidence that SD may be conditioned differently than *ne*-retention with respect to interlocutor differences. This discussion will be revisited in Chapter 6.

[90] These figures include all L2 French speakers: the 9 Near-NS participants and the two near-native interlocutors recruited for the data collection in Lille.

From an analysis of individual speakers, the significant difference for SD rates in Lille Near-NSs appears to be influenced primarily by a small subset of the nine speakers. Figure 5-12 provides the breakdown in SD rates by interlocutor type for these Near-NSs in Lille.

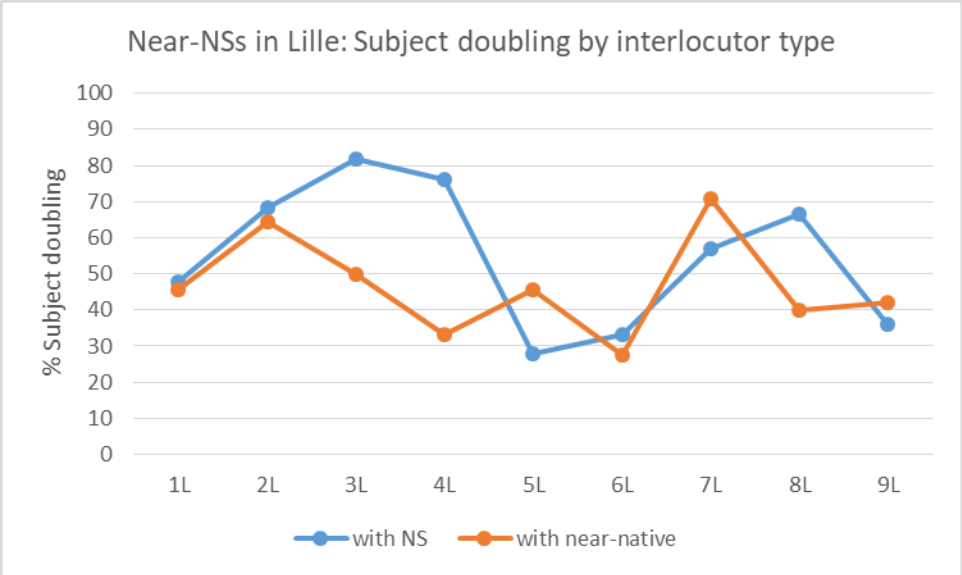

Figure 5-12. Subject doubling by interlocutor type for Near-NSs in Lille

For speakers 3L and 4L, SD rates with the NS interlocutor are 31.8% and 42.7% higher, respectively, than with the near-native interlocutor. These speakers do not share any immediately obvious personal characteristics (e.g., age, gender, length of residence) that might account for such a large difference across interlocutor type. Note, furthermore, that speaker 8L also shows a difference in SD (26.7%) nearly as large as speaker 3L, and in the same direction across interlocutor type. 8L does share some characteristics with 3L (younger age group, shorter length of residence), but other speakers sharing these characteristics (e.g., 1L, 3L, 9L) do not pattern in the same way. Regardless of these characteristics, it appears that the significant group difference is largely due to the individual variation of a small number of speakers, rather than a more uniform interlocutor effect observed in a majority of speakers, such as that found in the SA learner results.

By way of comparison, Figure 5-13 shows the SD rates by interlocutor type for the ten Near-NSs in Pau.

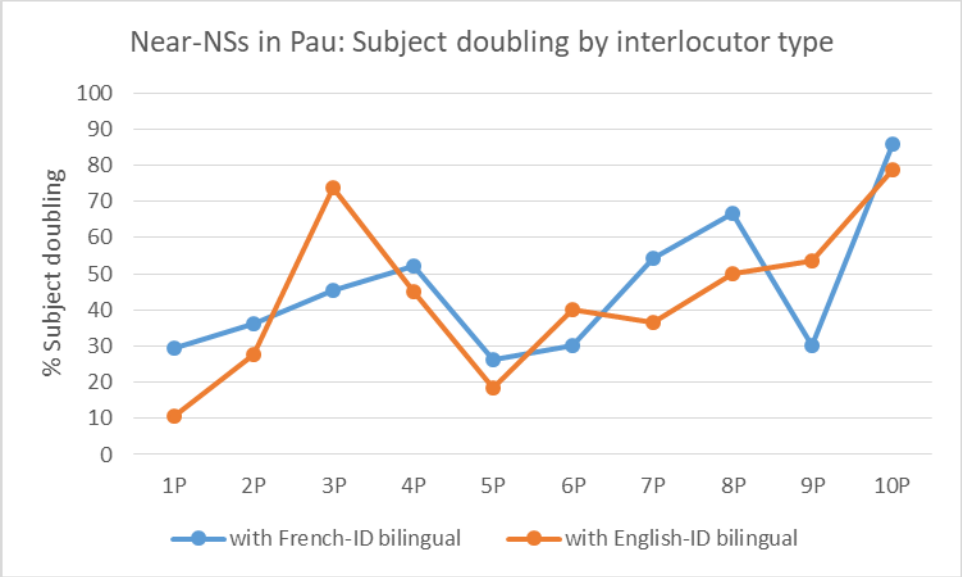

Figure 5-13. Subject doubling by interlocutor type for Near-NSs in Pau

As shown in the overall group percentages in Table 5-16, there was no significant difference in SD rates across interlocutor type for Near-NSs as a group in Pau. Compared with Lille, the percentages also differ less for individual speakers, with the highest differences observed in speaker 3P (28.2%) and 9P (23.3%), both of whom had higher SD rates with the English-identity bilingual interlocutor. It is interesting to note that only one other Pau speaker (6P) had higher SD rates with the English-identity interlocutor, though the actual differences across interlocutor type were marginal (less than 6%) for four other speakers. Furthermore, the aforementioned speakers 3P, 6P, and 9P do not all share immediately obvious personal characteristics that may explain a trend in one direction over another. Finally, when comparing the average difference in percentages across speakers (obtained by averaging the absolute value of the difference in percentage for each speaker at both sites), this difference is lower in Pau (14.5%) than in Lille (16.8%), suggesting less of an interlocutor effect with the bilingual speakers.

5.4.7 Strong pronouns

Here, I begin a discussion of certain linguistic features of SD that may help further explain native and non-native usage of this variable, beginning with strong pronouns. Recall from section 5.1.1 that third-person strong pronouns *il/elle/ils/elles* can occur without a coreferential subject clitic, but not first- and second-person ones such as *moi* and *toi.* Generally, these third-person strong pronouns are doubled by a corresponding subject clitic, but non-doubled structures are also attested in informal discourse. Nagy et al. (2003) and Zahler (2014) include third-person strong pronouns in their analysis, making cross-sample comparisons possible, though it must be noted that, since the Nagy et al. study concerned Montréal French, the distribution of strong pronouns is different from analyses of Hexagonal French. Furthermore, recall that feminine third-person *elle* was excluded from Zahler (2014) and in the current study as discussed in section 5.4.2.1. Looking only at *lui*, then, the rate of SD is 68% (40/59) in Nagy et al., similar to Zahler's (2014) overall strong pronoun rate of 70% SD (69/99).

Table 5-17 breaks down the use of strong pronouns *lui* and *eux* in the current study, where all participants (including all interlocutors) have been grouped together by L1 French status; this table does not include any data from SA learners, who produced no examples of strong pronoun *lui/eux* in subject position.

Table 5-17. Subject doubling with strong pronouns lui *and* eux

	lui	***eux***	**Total**
Near-NSs (all)	89.4% (42/47)	100% (24/24)	93.0% (66/71)
NSs (all)	88.9% (24/27)	85.0% (17/20)	87.2% (41/47)
Total	89.3% (67/75)	93.3% (42/45)	90.7% (107/118)

As Table 5-17 shows, SD frequency with strong pronouns in the current study is much higher than in the Nagy et al. (56%) and Zahler (70%) studies, and more similar to Fonseca-Greber's (2000:

313) results, who reported categorical SD except for one token of non-SD *eux*. There are only minor differences in the overall rates between the NS and Near-NS groups, suggesting that the tendency to favor SD with strong pronouns is fully acquired by speakers at this proficiency level in L2 French (and that the use of strong pronouns for discursive functions is widespread in high-proficiency learners). SA learners, however, demonstrate no evidence yet of acquiring this structure, resorting to emphasis on subject clitics or repetition of the corresponding lexical noun.

It is also worth noting that in the current study, strong pronouns account for only 6% of the non-clitic (i.e., lexical NP and strong pronoun) subjects in the current study, which is much lower than that obtained in Nagy et al. (16%) and Zahler (12%), though higher than in Ashby's (1980) Parisian study (2%). Strong pronouns in the current study therefore only slightly increase the overall SD average, from 52.1% (1191/2288) without to 53.9% (1298/2406) with strong pronouns.

5.4.8 Personal vs. neuter pronouns

Recall the data obtained from Ashby (1988) and Barnes (1985), described in section 5.2.3.1.1, which compared the preference for personal versus neuter pronouns in doubled contexts. Table 5-2 is reproduced in this section as Table 5-18.

Table 5-18. Personal versus neuter pronouns in SD contexts

Study	% SD	Doubled NPs / Total NPs	Neuter subject clitics / Total doubled NPs		Personal subject clitics / Total doubled NPs	
Ashby (1988)	73.3%	281/383	170/281	60.5%	111/281	39.5%
Barnes (1985)	79.2%	308/389	227/308	73.7%	81/308	26.3%

I compare these results from the Ashby and Barnes studies with the results for the current study, in Table 5-19, comparing each of the main speaker groups based on their use of personal and neuter

pronouns as doubling subject clitics. Column 1 lists the three main speaker groups. Column 2 provides the SD percentage in contexts where only a personal pronoun as doubling clitic would be possible, usually where the lexical NP has an obvious human referent (e.g., *Ma mère elle arrive demain* 'My mother (she) arrives tomorrow'). Column 3 indicates the SD percentages in contexts where only a neuter doubling clitic is possible (e.g., *Paris c'est beau,* 'Paris (it) is beautiful'). Column 4 includes the ambiguous contexts, where either a personal or a neuter doubling clitic is possible (e.g., *Votre français il est très bon / Votre français c'est très bon*, 'Your English (it) is very good'), and the SD percentages are broken down by pronoun type for each group. Finally, column 5 only includes contexts where SD occurred, reporting the distribution of personal and neuter doubling pronouns only in these contexts.

Table 5-19. Personal versus neuter pronouns in subject doubling contexts

	Personal pronoun contexts: % personal pronoun use (*il/s, elle/s*)		Neuter pronoun contexts: % neuter pronoun use (*ce, ça*)[91]		Ambiguous contexts: % personal & neuter use			% personal/ neuter in SD instances	
	%SD	Tokens	%SD	Tokens		%SD	Tokens		
SA learners	5.1	7/137	47.8	33/69	pers	0.5	1/210	8.2%	8/98
					neuter	26.7	56/210	91.8%	90/98
Near-NSs	34.6	223/645	77.6	184/237	pers	4.3	8/186	46.2%	227/491
					neuter	40.9	76/186	53.8%	264/491
NSs	55.0	378/687	85.8	211/246	pers	27.9	53/190	62.2%	429/690
					neuter	25.3	48/190	37.8%	261/690

From this table, we see a clear progression in the percentage of doubled personal pronouns used when a personal pronoun context is expected (column 2), increasing with proficiency. Furthermore, on the whole, all speakers double more when a neuter pronoun context appears

[91] No instances of impersonal *cela* doubling the subject NP were identified in any learners or interlocutors.

(column 3). In both personal and neuter contexts, the differences in SD rates between Near-NSs and NSs are statistically significant (personal: $\chi^2(1) = 56.2$; $p < .0001$; neuter: $\chi^2(1) = 5.36$; $p = .021$), though the difference in neuter contexts is much smaller. In columns 2 and 3, some interesting comparisons can be made with the L2 speakers in Nagy et al. (2003), who produced 29.0% SD (183/630) in personal pronoun contexts (similar to Near-NSs in the current study at 34.6% (223/645)), and who produced 85.7% SD (222/259) in neuter contexts, essentially indistinguishable from NSs in the current study at 85.8% (211/246). Another important result is observed in ambiguous contexts (column 4), where either a personal or a neuter doubling pronoun would be possible (subtle differences in meaning between personal/neuter doubling in those contexts notwithstanding): SA learners and Near-NSs overwhelmingly prefer neuter pronouns, whereas NSs slightly favor personal pronouns. This helps to explain the last column, where another clear progression can be seen: the higher the proficiency, the more personal pronouns are used as a percentage of all doubling clitic pronouns. We can compare these figures to those obtained in Barnes (1985) and Ashby (1988); see columns 4 and 5 in Table 5-18. The NSs in the current study behave almost identically to Ashby's NSs (for personal pronouns, 62% SD in the current study versus 60% in Ashby's NSs; for neuter pronouns, 38% in the current study versus 40% in Ashby's NSs).

SA learners produced eight SDs that involve personal pronouns (three *il*, one *ils*, three *elle*, one *elles*); in these cases, there were no errors matching the gender of the lexical noun with the gender of the clitic pronoun, suggesting that these learners were able to make online gender agreement where necessary, rather than defaulting to masculine *il/s*. However, in all but one of these cases, the lexical NP referred to a human, rendering this gender agreement comparatively

easy for learners.[92] The one instance of a non-human (e.g., [-animate]) lexical NP doubled by a personal subject clitic is as follows:

(50) *Mais tous les bouteilles de coca, ils sont en maximum 50 centilitres.* (3S)
'But all the bottles of Coke, they are maximum 50 centiliters.'

Though this speaker did not provide the correct grammatical gender for *bouteille* (feminine) in the quantifier *tous* (masculine), he did match the gender specification for *tous* with doubling *ils*. The gender specification in doubling clitics referencing [+animate] lexical NPs was also rather accurate in SA learners; there was only one instance of an SA learner using a neuter pronoun in a context requiring only a personal (not neuter) pronoun for doubling (according to NS judgments):

(51) *Oh non mon ami(e) c'est pas dans le même groupe*. (2S)
'Oh no my friend (it) is not in the same group.'

This may be an instance of substituting a lexicalized expression (*c'est*) lacking gender specification where an online gender specification would be required for a grammatical utterance. Furthermore, it is unclear whether the speaker had a specific friend in mind, since the sequence [monami] could have been referring to a female (as *mon ami*$_{\text{MASC}}$ and *mon amie*$_{\text{FEM}}$ are phonetically identical), further motivating a processing decision in which gender-unspecified *c'est* was inserted.

Regarding Near-NSs, where a grammatical gender decision needed to be made for doubling, all 8 examples were correctly marked for gender (e.g. *La température elle change pas*, 'The temperature (she) doesn't change'). For NSs, all 53 grammatical gender decisions were,

[92] These utterances are provided here:
Et ma mère elle est un peintre 'And my mother (she) is a painter' (2S)
Ma mère d'accueil elle m'a dit que il y a une forte communité. 'My host mother (she) told me that there is a strong community.' (3S)
Mon frère d'accueil ici il parle et c'est difficile. 'My host brother here (he) speaks and it's difficult.' (4S)
Les petites toujours elles parlent très vite. 'The little ones (they) always speak very quickly.' (5S)
Bernadotte il a devenu le roi de Suède ? 'Bernadotte (he) became the king of Sweden?' (5S)
Ma sœur d'accueil qui a 12 ans, elle était dans un spectacle. 'My host sister who's 12 years old, she was in a talent show.' (5S)
Mon père d'accueil il parle parfaitement l'espagnol. 'My host father (he) speaks Spanish perfectly.' (8S)

unsurprisingly, made correctly. As mentioned above, it is clear that, despite similar overall SD rates between Near-NSs (46%) and NSs (53%) in ambiguous contexts, Near-NSs strongly prefer neuter pronouns, while NSs slightly prefer personal pronouns. The fact that NSs are much more likely to double with personal pronouns when a grammatical gender decision must be made appears to be a robust difference between native and non-native speakers as a whole. This difference is likely due not only to the Near-NS overuse of lexicalized *c'est* but also the online gender agreement that must be made when doubling with a personal pronoun, which is more of a risk for these non-natives, who must remember features of the NP, such as grammatical gender, that are not present in their native language. A revealing comparison could, however, be made with L2 French speakers whose L1 contains grammatical gender, an analysis that may be explored in a future study (see Trévise (1986) for evidence of L1 Spanish speakers using more personal pronouns compared with L1 English speakers).

5.4.8.1 Further discussion of ambiguous contexts

For the current study, Table 5-20 lists all speakers who used at least one personal pronoun as subject clitic to double an NP in ambiguous contexts. Column 1 groups the individual speakers by their language status; column 2 provides the speaker ID for each speaker. Column 3 provides the tokens and percentages of SD with personal pronouns for all ambiguous contexts (i.e., contexts that present a gender computation challenge, since neuter pronouns lacking gender specification may be used felicitously instead; cf. non-doubled *Votre français est très bon* 'Your French is very good' versus doubled *Votre français il est très bon* 'Your French (he) is very good'). Column 4 considers only cases of SD produced in ambiguous contexts, reporting the percentage of doubling personal pronouns when SD occurred; the percentage of doubling neuter pronouns can therefore be obtained from the inverse of these percentages.

Table 5-20. Doubling personal pronoun usage in ambiguous contexts

Language status	Speaker ID	SD with personal pronouns out of all ambiguous contexts		% personal pronouns out of SD in ambiguous contexts	
SA learner	3S	1/16	6%	1/7	14%
Near-NS	8L	1/16	6%	1/6	33%
	2P	1/4	25%	1/2	50%
	3P	2/9	22%	2/6	33%
	4P	1/4	25%	1/4	25%
	7P	1/18	6%	1/7	14%
	8P	2/14	14%	2/8	25%
NS (w/ Near-NSs)	CaF	10/34	29%	10/20	50%
	KeF	13/31	42%	13/16	81%
	Ch	10/59	17%	10/33	30%
	Fr	12/48	25%	12/22	55%
	Th	8/18	44%	8/10	80%
Near-native (w/ SA learners)	AmE	1/13	7%	1/6	17%
NS (w/ SA learners)	SoF	2/15	13%	2/9	22%
All near-native		9/78	12%	9/39	23%
All NS		53/190	28%	53/101	53%

Of the six Near-NSs who produced personal pronouns in ambiguous contexts, there is some overlap with the SD rates of NSs, though the small number of tokens for each Near-NS precludes any definitive conclusions. It would be instructive to obtain a larger sample in order to determine whether the Near-NSs with the highest personal pronoun rates would pattern similarly with NSs;[93] it cannot be ruled out that some Near-NSs would show nativelike performance in a larger sample of this context of SD use. However, as a group, Near-NSs (and SA learners) show clear differences in SD distribution compared with NSs.

[93] Table 5-20 includes all non-native speakers who produced SD in ambiguous contexts, including the near-native interlocutor for SA learners (AmE). The near-native interlocutors in Lille produced 28 ambiguous contexts but no doubling personal pronouns.

5.4.8.2 Stages in the acquisition of SD

The results reported in this chapter point toward a sequence that can be observed in the development of variable SD in adult L2 French, as represented in Figure 5-14.

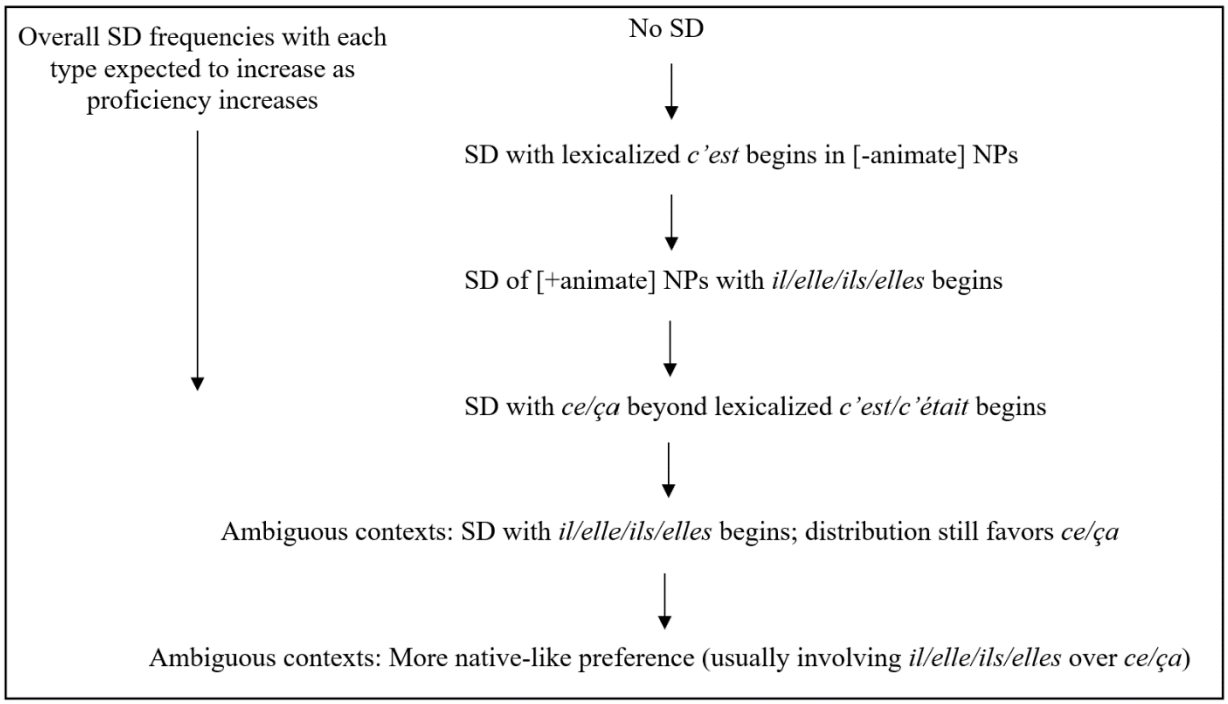

Figure 5-14. Stages in the acquisition of variable subject doubling

If we assume L1 parameters as the initial state for native English speakers of French, we would expect little or no SD use early in the acquisition process, especially for those learners exposed to the language in instructional settings, where written (i.e., standard) forms are likely to be more frequently encountered. The following stages then apply to learners exposed to spoken forms of French, which may include exposure to some SD in formal styles but whose input is overwhelmingly likely to include exposure to frequent SD in informal styles. Even then, as the data for SA learner 6S attest (no tokens of SD in 45 possible contexts), there may be a delay between exposure to SD in informal contexts and SD production in the learner's speech. As

learners recognize the utility of the lexicalized expression *c'est*, they may begin a stage of doubling with this expression (and possibly *c'était*) in [-animate] NPs (e.g., *Pau c'est joli*), which may also include infelicitous use (i.e., over-reliance) of neuter *ce* with [+animate] NPs (e.g., *mon ami(e) c'est pas dans le même groupe* 'my friend (it) is not in the same group'). Personal pronoun doubling of [+animate] NPs (e.g., *ma mère elle arrive demain* 'my mother (she) arrives tomorrow') then follows, as seen in the data of several SA learners in the current study, but at much lower frequencies than doubling with lexicalized *c'est.*[94] At more advanced levels, the frequency of SD in all of these contexts is likely to increase but remain lower than NS frequencies, while the use of doubling neuter *ce/ça* in non-lexicalized contexts (e.g., *la température ça change vite* 'the temperature (it) changes quickly') becomes more frequent.[95] At high proficiency levels, especially in learners demonstrating frequent interaction with, and integration into, target language communities, the use of doubling personal pronouns in ambiguous contexts may begin to appear, though such learners are still likely to over-rely on neuter *ce/ça* in such contexts. Again, this over-reliance may be due to avoidance in making overt gender agreement marking with [-animate] NPs (thus creating the possibility for gender agreement errors), or it may be due to L1 interference in attributing grammatical gender to [-animate] NPs. Highly proficient speakers may also reach another developmental stage in demonstrating more fine-grained sensitivity to the preference of either a personal or a neuter pronoun in putative ambiguous contexts where NSs also prefer one type of pronoun over the other. NSs often prefer personal pronouns in these contexts (e.g., *ma chambre elle est à côté* 'my room is nearby' is preferred over *ma chambre c'est à côté*). In other

[94] The data from the SA learner group also indicate that obligatory SD with strong pronoun *moi* (e.g., *Moi, je…* 'Me, I…') is likely to appear at this stage; three of the five learners producing SD with personal pronouns also produced the *moi, je…* structure. However, since the latter structure is not an example of variable SD, it is not included in Figure 5-14.

[95] In the current study, two tokens of this structure were produced in the SA learner group (by the same learner), but this structure was attested in nearly all Near-NSs.

cases, properties of the NP provide restrictions on the type of doubling pronoun. For example, subject specificity favors SD with personal pronouns over neuter pronouns (recall similar observations in section 5.2.3.1.1), as illustrated in (52).

(52) a. More specific: *Cet alcool il est très fort.* ('This alcohol (he) is very strong.')
b. More general: *L'alcool ça déshydrate.* ('Alcohol (it) dehydrates you.')

Semantic differences can also favor one doubling pronoun over another. In the expression *Pour moi la maison est en Angleterre* (lit., 'For me the house is in England'), the interpretation given by a native speaker rater of this sentence was that doubling with a personal pronoun (*Pour moi la maison elle est en Angleterre*) would suggest a more specific, concrete reference to the house belonging to the speaker, while doubling with a neuter pronoun (*Pour moi la maison c'est en Angleterre*) connotes the idea of the speaker being "chez moi" ('at home/at my home'). Given the context, and the introductory clause *pour moi*, the latter interpretation was judged to be preferable, which is indeed what the speaker (bilingual 'Ch') produced with the doubling neuter pronoun.

Examples of fine distinctions such as these are difficult to elicit in large numbers in spontaneous oral production, especially in learners, and more detailed analyses into these distinctions would benefit from additional measures such as judgment tasks involving preferences of doubling pronoun types, which lie beyond the scope of the current study. Differences in the overall frequency of SD can still paint a revealing picture of how well learners integrate this structure into their informal styles. What complicates observations based on frequencies is that, in a given conversation, a speaker may produce structures containing more contexts favoring doubling in general, or doubling only with personal pronouns (or conversely, with neuter pronouns), and the resulting SD distribution may not be entirely accurate for modeling trends in SD by proficiency level as outlined in this section. However, just as syntactic, phonological, and semantic constraints condition the variable use of *ne*-retention, similar constraints may apply not

only to condition whether a given speaker is likely to produce SD in a given utterance, but also to condition whether the potential doubling clitic is personal or neuter. A variationist analysis such as those undertaken in previous studies on SD can account for the variation inherent at multiple levels of discourse, including not only structural constraints but also the topics of conversation, as well as, in the current study, potential interlocutor effects such as those found in Chapter 4. The following section describes the variationist analysis undertaken to further tease apart such effects on SD.

5.5 Variationist analysis: Subject doubling

As with the *ne*-retention variable, subject doubling can be expected to be influenced by multiple and simultaneous factors on the contexts in which SD may occur, as well as the nature of the potential doubling clitic. This section outlines the linguistic, extralinguistic, and sociostylistic factors that were used in the variationist analysis on SD in the current study, with explanations and examples of each.

5.5.1 Linguistic factors

Most of the linguistic factors outlined here have been identified in previous SD studies such as Nadasdi (1995b), Nagy et al. (2003) and Zahler (2014). Nagy et al. classify the linguistic factors into four main categories: subject type, verbal syntax, preverbal material, and semantic properties of the subject. In the current study I organize these categories somewhat differently, focusing first on clitic properties and lexical subject properties, followed by verbal syntax and then preverbal material. For each factor group, I describe the factors that were coded, with examples of each factor and discussion where appropriate.

5.5.1.1 Clitic properties

1) Clitic context: Lexical subjects and strong pronoun subjects can be doubled by only a personal pronoun, by only a neuter pronoun, or by either type (ambiguous).

Personal	*Ma mère elle parle allemand.*
	'My mother (she) speaks German.'
	**Ma mère ça parle allemand.*
	'My mother (it) speaks German.'
Neuter	*Être chercheur c'est difficile.*
	'Being a researcher (it) is difficult.'
	**Être chercheur il est difficile.*
	'Being a researcher (he/it) is difficult.'
Ambiguous	*La ville elle est sympa.*
	'The town (she) is nice.'
	La ville c'est sympa.
	'The town (it) is nice.'

As seen in section 5.4.8, speakers broadly double with personal pronouns in personal contexts and neuter pronouns in neuter contexts, but not always. Furthermore, native speakers may not always agree on whether a particular type of pronoun may be possible for a specific context—though learners may not necessarily follow the same context constraints as native speakers when choosing the type of doubling clitic. In their analysis, Nagy et al. (2003) note that ambiguous contexts allowing for either a personal or neuter pronoun were coded as such by the judgment of a native speaker. For my analysis, I use the judgments of the two native speaker raters recruited specifically for determining the contexts of variation for this variable, as discussed in section 5.4.1.

Note that the choice of clitic as personal or neuter is constrained both by properties of the subject as well as the syntax of the VP. Generally, lexical subjects referring to specific people can only be doubled by a personal pronoun, but in certain syntactic structures the same lexical subject can only be doubled by a neuter pronoun, as in (53c).

(53)	a. *Ma mère elle est intelligente.*	'My mother (she) is intelligent.'
	b. ??*Ma mère c'est intelligent*.	'My mother (it) is intelligent.'
	c. *Ma mère c'est une femme intelligente.*	'My mother (it) is an intelligent woman.'

d. **Ma mère elle est une femme intelligente.* 'My mother (she) is an intelligent woman.'

Judgments from my native speaker raters largely corresponded to these syntactic constraints.

5.5.1.2 Lexical subject properties

In cases where multiple lexical nouns appear in the subject NP, I coded for the noun subject nearest to the verb. Thus, in (54), *les étrangères qui viennent en Australie* was coded as the subject NP, with *les étrangères* coded as a common noun; for the purposes of coding for lexical subject properties, *les Australiens* was ignored.

(54) *Même les Australiens et aussi les étrangères qui viennent en Australie, ils trouvent que ça vaut pas vraiment le coup.*
'Even Australians and also foreigners who come to Australia, they find that it's not really worth it.'

The following factors correspond to the lexical subject properties coded in the variationist analysis.

2) Subject type: The subject can be a third-person strong pronoun (*lui/eux*), proper noun, common noun, indefinite pronoun, or a different syntactic category (verb or prepositional phrase).

Strong pronoun	*Lui il est arrivé.* 'Him (he) has arrived.'
Proper noun	*Chloé est arrivée.* 'Chloé has arrived.'
Common noun	*La dame est arrivée.* 'The lady has arrived.'
Other pronouns	*Certains sont arrivés.* 'Some have arrived.'
Verb/PP/other	*Être bilingue ça aide beaucoup.* 'Being bilingual (that) helps a lot.'

As discussed in section 5.4.2, the feminine strong pronouns *elle/elles* were not included for analysis. Other pronouns include demonstratives such as *celui-là* and possessives such as *le mien*, as well as the indefinite pronouns, including *autre, chacun, certains, beaucoup, quelqu'un, quelque chose, plusieurs*, *la plupart* (and the feminine and/or plural inflections of

these pronouns where appropriate). Due to their different structures, clauses beginning with relatives *ce qui/ce que* are included in the Verb/PP/other category (cf. Auger & Villeneuve, 2010: 72).

3) Subject definiteness: The subject can be definite, indefinite, quantified, or a Verb/PP.

Definite	*La table est ronde.* 'The table is round.'
Indefinite	*Des tables sont rondes.* 'Tables are round.'
Quantified	*Plusieurs sont rondes.* 'Several are round.'
Verb/PP	*Acheter une table ronde c'est difficile*. 'Buying a round table is difficult.'

Quantified subjects include the indefinite pronouns, but only when they are combined with lexical nouns, e.g. *la plupart de mes amis* 'most of my friends'; when they occur alone (e.g., *la plupart*) they are categorized as definites (Nadasdi, 1995b: 94). *Tout* + NP is also included in the quantified category; as discussed in section 5.4.2.2, bare *tout* and the expression *tout le monde* in subject position were excluded from analysis.

4) Subject specificity: The subject can be specific, non-specific, or generalizing.

Specific	*La personne a acheté cette table.* 'The person bought this table.'
Non-specific	*Une personne a acheté cette table.* 'A person bought this table.'
Generalizing	*Les gens sont fâchés contre le gouvernement.* 'People are angry with the government.'

As mentioned in section 5.2.3.1.2, Subject definiteness and specificity were combined in previous studies such as Nadasdi (1995b) but analyzed separately in more recent work (Auger & Villeneuve, 2010; Nagy et al., 2003).

5) Subject animacy: The subject can refer to an animate noun, a material but inanimate noun, an immaterial and inanimate noun, or a place.

Animate	*Ma mère elle parle allemand.* 'My mother (she) speaks German.'
Material but inanimate	*La table elle est ronde.* 'The table (she) is round.'
Immaterial and inanimate	*L'idée c'est de travailler en Angleterre.* 'The idea (it) is to work in England.'
Place	*Paris c'est beau.* 'Paris (it) is beautiful.'

Nagy et al. (2003) added "Place," derived from Fonseca-Greber (2000), in order to differentiate from other immaterial and inanimate subjects, though no differences between these two factors were found in that study.

5.5.1.3 Verbal syntax

6) Verb type: Is the verb of the doubled subject a transitive, intransitive, passive, modal, or copula? As with *ne*-retention, modal verbs include *aller* ('go'), *pouvoir* ('be able'), *vouloir* ('want'), *devoir* ('must'), and *falloir* ('must').[96]

7) Clause type: Is the clause containing the doubled subject a matrix (main), subordinate, or relative clause, or a conditional clause introduced by *si* ('if')?

Matrix	*Ma mère elle parle allemand.* 'My mother (she) speaks German.'
Subordinate	*Je pense que ma mère elle parle allemand.* 'I think my mother (she) speaks German.'
Relative	*La fille que ma mère elle connaît parle allemand.* 'The girl whom my mother (she) knows speaks German.'
si clause	*Je ne sais pas si ma mère elle parle allemand.* 'I don't know if my mother (she) speaks German.'

[96] Zahler's (2014) all-NS study found a favoring effect for SD with certain high-frequency verbs (non-modal *aller, avoir, faire,* and *pouvoir*), grouping these verbs together as a separate factor for verb type. However, in the current study, SD usage with these verbs (n = 258) was nearly identical to the overall SD average for all groups except for SA learners, who had categorical absence of SD with these verbs (n = 24) compared with 24% SD overall. The data from the current study suggest that this effect of frequency may not be particularly robust. I did not include this factor in the variationist analysis.

As per Nagy et al. (2003), generally, the matrix clause subject begins an utterance, or it may be introduced by *et* ('and'), *ou* ('or'), or *mais* ('but'). Subordinate clauses begin with *parce que* ('because'), *quand, lorsque* ('when'), *comment* ('how'), etc. Relative clauses begin with the relative pronouns *qui* ('who'), *que* ('that'), *dont* ('of which'), *où* ('where'), and variants thereof. Since Nagy et al. (2003) found that subordinate clauses introduced by *si* contained much lower SD frequencies (11.8%, 2/17) than other subordinate clauses (38.5%, 42/109), and since Auger and Villeneuve (2010) found a similar trend (though a smaller difference) in NSs, this type of clause will be analyzed separately to see if this distinction holds in the current study.

5.5.1.4 Preverbal material

These factors generally follow the specifications in Nagy et al. (2003).

8) Relative clause: The lexical subject may be followed by a relative clause modifier. If so, the matrix clause verb, and any doubled subject, will follow the relative clause.

Relative clause	*La fille qui est partie elle parle allemand.*
	'The girl who left (she) speaks German.'
No relative clause	*La fille elle parle allemand.*
	'The girl (she) speaks German.'

9) Preverbal clitics: Object clitics (including reflexives) may appear between the lexical subject and the verb.

Object clitic	*Ma mère elle me téléphone.*	'My mother (she) calls me.'
Reflexive clitic	*Ma mère elle s'habille en blanc.*	'My mother (she) wears white.'
No clitic	*Ma mère elle parle allemand.*	'My mother (she) speaks German.

Following Nagy et al. (2003), the preverbal clitic *ne* was analyzed in a separate factor group.

10) Negation: In verbal negation contexts with the lexical subject, *ne* may be retained or deleted.

ne-retention	*Ma mère elle ne parle pas allemand.*

	'My mother (she) NEG doesn't speak German.'
ne-deletion	*Ma mère elle parle pas allemand.*
	'My mother (she) doesn't speak German.'
non-negative context	*Ma mère elle parle allemand.*
	'My mother (she) speaks German.'

I divide this factor group into three categories, per Zahler (2014).

11) Other preverbal material: In addition to preverbal clitics and *ne,* there may be other intervening material between the subject and the verb.

Adverb	*La fille ici elle parle allemand.*
	'The girl here (she) speaks German.'
Hesitation	*La fille…elle parle allemand.*
	'The girl…(she) speaks German.'
Parenthetical	*La fille par exemple elle parle allemand.*
	'The girl, for example, (she) speaks German.'
Prepositional phrase	*La fille sur la photo elle parle allemand.*
	'The girl in the picture (she) speaks German.'
Feedback	*La fille* [*oui bien sûr*] *elle parle allemand.*
	'The girl [yes of course] (she) speaks German.'
Multiple elements	*La fille sur la photo…elle parle allemand.*
	'The girl in the picture…(she) speaks German.'
No material	*La fille elle parle allemand.*
	'The girl (she) speaks German.'

Note that a PP complement (e.g., *La fille <u>sur la photo</u>*) produces a complex NP subject and may be preferentially grouped with NPs modified by relative clauses (as in factor #8) rather than with other preverbal material. To allow for more direct comparisons with the Nagy et al. (2003) study, I maintain their coding scheme regarding this structure. An additional comment is warranted in the coding for "hesitation," as what constitutes hesitation has not been clearly defined in previous literature. In the current study, I counted as hesitation contexts where the speaker either produced a filler such as *euh* or (especially with SA learners) *um*, or the speaker paused for a minimum of about one second before producing the verb. Multiple elements were coded whenever more than one of the preverbal elements in this category appeared (e.g., hesitation *and* prepositional phrase), per Zahler (2014).

5.5.2 Extralinguistic factors

Extralinguistic factors considered for the variationist analysis of SD were the same as for the *ne*-retention analysis (see sections 4.5.3 and 4.5.4 for details on these factors). For SA learners, these factors were Gender (male, female), Interlocutor L1 (English, French), Interlocutor type (NS, Near-NS, SA learner), Conversation portion (0-5 minutes, 5-10 minutes, 10+ minutes), with SpeakerID as a random intercept. For Near-NSs, these factors were Interlocutor L1 (Pau group: English identity/French identity; Lille group: L1 English/L1 French), Age (younger (-38), older (44+)), Gender, and Conversation portion, with SpeakerID as a random intercept. For the NSs, the factors include Conversation portion and, for NSs in Lille, English or French identity, with SpeakerID as a random intercept.

5.5.3 Sociostylistic factors

Compared with *ne*-retention, sociostylistic factors for the variationist analysis of subject doubling were limited to the *tu/vous* category. Concerning quoted speech, of the utterances in which quoted speech occurred in possible SD contexts over all speakers, there were 15 tokens which could be presumed to be in a formal style; of these tokens, 3 of 10 were doubled in Near-NSs and 1 of 5 in NSs. Due to such low token counts, this factor was not included for analysis. Furthermore, since the category of "emphasis" is addressed differently with regard to SD compared to *ne*-retention, it was not included here either.

The topics of conversation were coded to determine if serious topics led to use of the more formal variant (absence of SD). However, in all of the speaker groups, there is no decrease in SD frequency for topics coded as serious (see Table 5-21).

Table 5-21. SD rates in serious topics for each speaker group

Speaker group	SD tokens in serious topics	% SD in serious topics	%SD overall
SA learners	25/63	39.7	23.6
SA interlocutors	26/31	83.9	66.0
Near-NSs	78/154	50.6	46.0
NS interlocutors	158/253	62.5	61.3
Lille near-native interlocutors	24/40	60.0	54.4

Again, the selection of "serious" topics can be somewhat problematic for many of these speakers, as it can be difficult to determine which utterances fall within such broad topics as "education" or, in the case of SA learners, whether any utterances can be considered under the topic of "one's profession." The results in the above table must therefore be interpreted cautiously, and due to the reasons cited here, this factor group was not included in the analysis.

5.5.4 Results: SA learners

All participants were coded for the factor groups as described earlier in this section, with the presence or absence of doubling subject clitic as the dependent variable, and Rbrul (Johnson, 2009) was used to carry out a variationist analysis. One learner (6S) was excluded from the Rbrul analysis due to categorical absence of SD in all three of his conversations (comprising 45 possible SD contexts), resulting in 26.4% SD (98/371) in SA learners for the tokens analyzed through Rbrul. As with the *ne*-retention data, the factor "SpeakerID" was run as a random intercept. These data thus report on the 7 remaining SA learners, over 23 conversations. Table 5-22 shows the significant factor groups for SA learners, ranked in order of significance. The non-significant factor groups are listed below the table; see Appendix G for details on these non-significant factors. Recall that in all subsequent tables reporting significant factor groups, these groups are ranked by order of significance.

For SA learners, due to low token counts, several factors were regrouped after an initial Rbrul run. In the Verb type category, Passives (7 tokens; no SD) were combined with Modals; in the Clause type category, Relatives (6 tokens; no SD) and *si* clauses (3 tokens; no SD) were grouped with Subordinates; in the Pre-verbal material category, Feedback (2 tokens; no SD) was grouped with Hesitations, and Parentheticals (3 tokens; 1 SD) were grouped with Prepositions; in the Subject definiteness category, Indefinites (4 tokens; 1SD) were grouped with Quantifiers, and Verbs (2 tokens; 1 SD) were grouped with Definites.

Table 5-22. Significant factor groups for SA learners (subject doubling)

Factor Group	**Factor**	**Log odds**	**Factor weight**	**% SD**	**N**
Verb type	copula	5.705	.997	38.6	91/236
	intransitive	4.456	.989	14.3	5/35
	transitive	2.489	.923	2.5	2/79
	modal/passive	-12.650	.001	0.0	0/21
Clause type	matrix	0.802	.690	31.4	92/293
	subordinate/relative/*si*	-0.802	.310	7.7	6/78
Clitic context	neuter	0.929	.717	52.4	33/63
	ambiguous	-0.278	.431	30.8	57/185
	personal	-0.651	.343	6.5	8/123
Other pre-verbal material	multiple elements	1.284	.783	12.5	2/16
	PP	1.034	.738	26.5	18/68
	adverb	0.014	.503	21.4	3/14
	none	-0.488	.380	29.5	71/241
	hesitation	-1.843	.137	12.5	4/32
Subject definiteness	definite/verb	0.892	.709	2.8	95/330
	quantifier/indefinite	-0.892	.291	7.5	3/41
Interlocutor status	NS	0.503	.623	31.8	34/107
	Near-NS	0.223	.556	32.7	53/162
	SA learner	-0.727	.326	10.8	11/102
TOTAL				26.4%	98/371

Input probability = 0.264; Log likelihood = -132.415
Non-significant factor groups (cf. Table G-10): Subject type, Subject specificity, Subject animacy, Relative clause, Negation, Preverbal clitics, *Tu/vous*, Conversation portion, Gender, Interlocutor L1

As the results from this variationist analysis indicate, there is a primacy of linguistic factors over non-linguistic factors (cf. Preston, 1991, and discussion in section 2.10). Crucially, as with

ne-retention, the interlocutor's language status appears as a significant factor group (and the only non-linguistic factor group) for subject doubling, with SD slightly favored in conversation with the NS and near-native, and SD more strongly disfavored (as evidenced by the comparatively lower log-odds) with other SA learners. The presence of this significant factor group supports the hypothesis that interlocutor language background influences multiple sociolinguistic variables, at least in this learner proficiency group.

The significant linguistic factor groups for SA learners mostly overlap with the L2 speakers in Nagy et al. (2003), with four of the five factor groups appearing in Nagy et al.'s study—and these four in the same order of significance. However, the outstanding difference is in the most significant factor group in SA learners, Verb type, which is not significant in Nagy et al. (but significant in the L1 speakers to which their learners were compared). Conversely, the Subject type factor group was most significant in Nagy et al. but not selected in the current analysis. This is despite the fact that even the SA learners largely follow the same hierarchy within this factor group that most previous studies (e.g., Nadasdi, 1995b) have identified (see Table 5-23):

Table 5-23. SD rates according to subject type in SA learners (all)

Noun Type	**Tokens**	**% SD**
1st- & 2nd-person strong pronouns[97]	19/20	95
3rd-person strong pronouns	--	--
Proper noun	27/75	36.0
Common noun	70/282	24.8
Indefinite pronoun/all other pronouns	0/12	0.0

[97] Only *moi* and *toi* are included in these data as strong pronouns clearly discernable from their co-referential subject clitics. The lone example of a non-doubled strong pronoun was in the ungrammatical utterance by SA speaker 7S: *et toi reviens de Nebraska?* ('and you come back from Nebraska?'). Note also that only one other instance of *toi* in subject position was produced; the remaining 18 tokens were of the *moi je…* variety.

The main difference between these two studies is, of course, that SA learners produced no third-person strong pronouns in subject position. This may partially explain the large difference in significance across the two learner populations. As for Verb type, SA learners and Nagy et al.'s learners both trend in the same direction regarding copulas, which most strongly favor SD; however, SA learners strongly disfavor SD with all other verb types, while more advanced learners in Nagy et al. only slightly disfavor SD with all other verb types. As discussed in section 5.4.8.2, SA learners seem to have identified the copula (generally in the form of the expression *c'est* and sometimes in the imperfect form *c'était*) as a common environment for doubling, almost exclusively with the neuter pronoun *ce*. At this stage in their sociolinguistic development, *c'est*, in addition to being a lexicalized expression with minimal phonological complexity, also requires no computations of gender agreement and can override number agreement with the lexical NP. It is perhaps unsurprising that this factor group is the most significant for this learner group.

Regarding the other significant verbal factor for SA learners, Clause type, SD is favored in matrix clauses over subordinate clauses overall, as in Nagy et al.; however, the difference in clause type for SA learners is much greater (31% SD in matrix clauses, 8% in all subordinate clauses) than in Nagy et al. (47% in matrix clauses, 35% in all subordinate clauses). This larger difference in SA learners may be due to the greater syntactical complexity in subordinate clauses having a stronger inhibiting effect on the use of informal variants in lower-proficiency learners.

Though negation was not selected as significant, SA learners treat it in the same way as in previous studies: SD is categorically disfavored in *ne*-retention (0%) but strongly favored in *ne*-deletion (73%). The small token counts ($n = 27$) in negation contexts may explain why this factor group was not significant. This context will be discussed further in Chapter 6.

Regarding non-significant extralinguistic factors, it is perhaps surprising that in conversations where SA learners used *vous* as the address pronoun, they had higher SD rates than in conversations using *tu* (38% SD in 56 tokens with *vous* versus 26% SD in 309 tokens with *tu*). One learner in particular (4S) may account for this result, as she produced SD with 17 of 25 tokens (68%) in *vous* contexts; excluding this learner yields 13% SD in *vous* contexts and 22% SD in *tu* contexts. As for the conversation portion, there is a slight tendency toward lower SD rates as the conversation progressed, suggesting that learners were not affected by the recording environment, at least as far as the production of the informal variant of this sociolinguistic variable is concerned.

As in Nagy et al. (2003), Subject animacy, Subject relative clauses, and Pre-verbal clitics were not selected as significant. Broadly, the fact that there is a large overlap between the significant and non-significant factor groups across both studies suggests that even at the proficiency level of these SA learners, and despite their lower overall SD frequency (27% versus 45%), the grammar of these learners patterns similarly to more advanced L2 speakers.

5.5.4.1 Results: Interlocutors with SA learners

Table 5-24 provides the significant factor groups for the native and near-native interlocutor with the SA learners. These data involve the two speakers in 14 conversations overall. Note that for the category of pre-verbal clitics, due to low token counts, reflexive pronouns and object clitics were grouped together.

Table 5-24. Significant factor groups for SA interlocutors (subject doubling)

Factor Group	Factor	Log odds	Factor weight	% SD	N
Subject type	strong pronoun	15.049	> .999	100.0	4/4
	indefinite/other pronoun	-3.457	.031	77.8	7/9
	proper noun	-4.525	.011	87.5	35/40
	common noun	-7.067	.001	53.4	47/88
Clitic context[98]	neuter	2.802	.943	95.7	44/46
	ambiguous	-1.242	.224	50.0	13/26
	personal	-1.560	.174	52.2	36/69
Subject animacy[99]	inanimate & material	15.802	> .999	100.0	7/7
	animate	-4.259	.014	53.2	33/62
	inanimate & immaterial	-4.899	.007	63.2	24/38
	place	-6.644	.001	85.3	29/34
Pre-verbal clitics	none	1.633	.837	67.7	90/133
	reflexive/object clitic	-1.633	.163	37.5	3/8
Subject specificity[100]	generalizing	0.935	.718	55.2	16/29
	specific	0.812	.693	71.4	70/98
	non-specific	-1.747	.148	50.0	7/14
Conversation portion[101]	0-5 min	1.165	.762	86.7	26/30
	5-10 min	-0.197	.451	66.7	12/18
	10+ min	-0.968	.275	59.1	55/93
Relative clause[102]	relative clause	1.347	.794	90.0	9/10
	no relative clause	-1.347	.206	64.1	84/131
TOTAL				66.0%	93/141

Input probability = 0.660; Log likelihood = -50.104
Non-significant factor groups (cf. Table G-11): Clause type, Verb type, Other pre-verbal material, Negation, Subject definiteness

This variationist analysis of the SA interlocutors may be less generalizable to analyses of other speaker groups due to smaller token counts, with only two speakers. Subject specificity, Conversation portion, and Relative clause were not selected in the Step-up model, so the significance of these factor groups is less certain. Nevertheless, comparisons can be made with

[98] The VIF (variance inflation factor) for this factor group was above 7.5, suggesting collinearity with another factor. Recall that values above 5 are thought to show that a predictor is highly correlated with the others.
[99] The VIF for this factor group was above 7.5.
[100] This factor group was not selected in the Step-up model.
[101] This factor group was not selected in the Step-up model.
[102] This factor group was not selected in the Step-up model.

previous studies. For Subject type, the "subject pronoun-proper noun-common noun" hierarchy observed in Nadasdi (1995b) and in Auger and Villeneuve (2010) obtains in these speakers as well. Note that the relatively high SD rate (78%, 7/9) for the indefinite/other pronoun category appears surprising, though the inclusion of "other" pronouns (such as *celui* 'the one' and *ceux* 'those') accounts for most of the SD use in these nine tokens. The results for the type of possible doubling clitic, where contexts for a neuter clitic favor SD over personal or ambiguous contexts, are similar to those observed in Nagy et al. (2003). Subject animacy also aligns with Nagy et al., where inanimate subjects and subjects of place favor SD over animate subjects. Finally, even though there are few tokens of pre-verbal clitics, there is still a significant inhibiting effect of these clitics on SD, as noted in Auger and Villeneuve (2010) and, to a lesser extent, in Nadasdi (1995b).

Furthermore, even for the factor groups identified as non-significant in this variationist analysis, the SD percentages still demonstrate trends in the same direction as observed in previous studies: slight favoring of SD in matrix over subordinate clauses, favoring of SD when other pre-verbal material appears, and favoring of copulas and transitive verbs over intransitives and modals. The low token counts may explain the non-significance of these groups. Nevertheless, the results from this smaller speaker group are still representative of general trends obtained in the other speaker groups in the current analysis; as we will see in the following sections, the two most significant factor groups for SA interlocutors—Clitic context and Subject type—will prove to be the two most significant groups from here on.

5.5.5 Results: Near-NS group and interlocutors

As with *ne*-retention, separate Rbrul runs were conducted for the Near-NS and NS groups at each site, followed by a run combining the Near-NSs at both sites and a run combining the NSs

at both sites. Table 5-25 provides the significant factor groups for the Near-NSs in Lille. These results are drawn from the 18 conversations involving the 9 Near-NSs in Lille.

Table 5-25. Significant factor groups for Near-NSs in Lille (subject doubling)

Factor Group	Factor	Log odds	Factor weight	% SD	N
Clitic context	neuter	2.194	.900	91.2	83/91
	ambiguous	-0.425	.395	45.5	35/77
	personal	-1.768	.146	40.0	86/215
Subject type	verb/PP	12.624	> .999	100.0	10/10
	strong pronoun	-0.429	.394	96.0	24/25
	proper noun	-3.783	.022	49.0	24/49
	common noun	-4.119	.016	50.2	132/263
	indefinite/quantifier	-4.293	.013	38.9	14/36
Clause type	matrix	5.105	.994	59.3	169/285
	subordinate	4.521	.989	40.2	33/82
	relative	3.248	.963	25.0	2/8
	si	-12.784	< .001	0.0	0/8
Subject animacy[103]	animate	1.197	.768	46.8	88/188
	inanimate & immaterial	-0.163	.459	64.2	86/134
	inanimate & material	-0.474	.384	36.4	4/11
	place	-0.560	.364	52.0	26/50
Verb type	copula	0.824	.695	65.8	125/190
	intransitive	0.390	.596	46.5	40/86
	modal	0.263	.565	48.0	12/25
	transitive	-0.215	.447	33.8	26/77
	passive	-1.263	.221	20.0	1/5
TOTAL				53.3%	204/383

Input probability = 0.246; Log likelihood = -182.3
Non-significant factor groups (cf. Table G-12): Other pre-verbal material, Subject specificity, Relative clause, Subject definiteness, Negation, Preverbal clitics, Conversation portion, Gender, Interlocutor L1, Age, Length of residence

As for the Near-NSs in Pau, Table 5-26 provides details on the significant factor groups (ranked according to significance). These results come from the 20 conversations involving the Near-NSs in Pau.

[103] The VIF for this factor group was above 2.5, suggesting that these values may be correlated with another predictor.

Table 5-26. Significant factor groups for Near-NSs in Pau (subject doubling)

Factor Group	Factor	Log odds	Factor weight	% SD	N
Clitic context[104]	neuter	1.231	.774	69.0	100/145
	ambiguous	-0.029	.493	40.0	40/100
	personal	-1.202	.231	33.4	147/440
Subject type[105]	strong pronoun	2.169	.897	89.2	33/37
	verb/PP	0.058	.515	88.2	15/17
	proper noun	-0.526	.371	34.5	39/113
	common noun	-0.832	.303	38.7	180/465
	indefinite/quantifier	-0.869	.295	37.7	20/53
Negation	*ne*-deletion	1.419	.805	73.7	28/38
	affirmative	0.328	.581	42.4	255/602
	ne-retention	-1.748	.148	8.9	4/45
Relative clause	relative clause	0.710	.670	68.5	37/54
	no relative clause	-0.710	.330	39.6	250/631
Subject animacy[106]	animate	0.762	.682	39.7	148/373
	inanimate & material	0.441	.609	47.8	11/23
	inanimate & immaterial	-0.290	.428	48.2	109/226
	place	-0.913	.286	30.2	19/63
Clause type	matrix	0.969	.725	46.8	237/506
	subordinate	0.311	.577	30.3	47/155
	relative	-0.469	.385	10.0	1/10
	si	-0.811	.308	14.3	2/14
Subject specificity	specific	0.488	.620	48.1	228/474
	generalizing	-0.089	.478	24.6	28/114
	non-specific	-0.399	.402	32.0	31/97
Verb type	copula	0.669	.661	51.7	156/302
	transitive	0.256	.564	36.6	71/194
	intransitive	0.017	.504	33.3	44/132
	modal	-0.240	.440	32.4	12/37
	passive	-0.703	.331	20.0	4/20
TOTAL				41.9%	387/685

Input probability = 0.378; Log likelihood = -343.615

Non-significant factor groups (cf. Table G-13): Subject type, Subject definiteness, Preverbal clitics, *Tu/vous*, Conversation portion, Gender, Age, Length of residence, Interlocutor identity

[104] The VIF for this factor group was above 2.5.

[105] The VIF for this factor group was above 2.5.

[106] The VIF for this factor group was above 5, suggesting that these values are correlated with another predictor.

As the two tables for these Near-NSs show, initial broad observations indicate that the Near-NSs at both sites share similar significant factor groups, suggesting that these speakers overall are largely influenced by the same factors conditioning SD usage, and that social factors have relatively little influence on SD rates.

The two most significant factor groups for each of these Near-NS groups, as with the SA interlocutor group, are Clitic context and Subject type. For Clitic context, the factors obtain as in Nagy et al.'s L2 speakers and in Nadasdi's NSs, with the current study's Near-NSs favoring SD when the doubling clitic must be neuter *ce/ça* rather than when doubling with a personal pronoun or either pronoun type is possible. For Subject type, Near-NSs at both sites almost always double strong pronouns, though both groups favor SD slightly more with common nouns than with proper nouns, in contrast with previous studies. For Clause type, relatives and *si* inhibit SD more in Near-NSs (13%) compared to other subordinate clauses (30%), as was found in studies on NSs, though small token counts (n = 24) require this observation to be made cautiously. Negation contexts aligned with observations in previous studies, with *ne*-deletion favoring SD and *ne*-retention inhibiting SD, though this effect was stronger in Pau than in Lille, likely due to the paucity of SD contexts occurring with *ne*-retention in Lille (n = 11). Subject animacy is broadly similar across sites, with inanimate subjects favoring SD over animate subjects (as observed in Auger & Villeneuve, but not in Nadasdi), though the relatively high SD rates for subjects of place observed in the SA interlocutors do not hold for Near-NSs in Pau.

For a visual summary of how the significant factors compare at both sites, Table 5-27 provides the rankings of the significant factor groups for Near-NSs in Lille and Pau.

Table 5-27. Significant factor groups for Near-NSs in both sites (subject doubling)

Near-NSs in Lille	**Near-NSs in Pau**
Clitic context	Clitic context
Subject type	Subject type
Clause type	Negation
Subject animacy	Relative clause
Verb type	Subject animacy
	Clause type
	Subject specificity
	Verb type

As this table shows, all significant factor groups in Lille were significant in Pau. The two most significant factors are the same (Clitic context and Subject type), and the remaining rankings for Lille are in the same order in Pau, except that Subject animacy and Clause type were flipped. The major exceptions are that certain pre-verbal material significantly influences SD variation in Pau. It is true that negation and relative clause contexts appear, as a percentage of overall tokens, slightly more often in Pau than in Lille (12.1% versus 11.2% for negation and 7.9% versus 6.3% for relative clauses), but these differences are too small to account for the significance of these factor groups. One explanation may be that these factor groups are significant when the threshold of a total number of tokens is reached, which may be the case in Pau but not in Lille.

No extralinguistic/sociostylistic factors were significant for either speaker group. As discussed earlier, the differences in Lille between interlocutor types were significant in a *chi*-square test but not significant in Rbrul. Age differences were more pronounced in Pau than in Lille, and there is a similar difference for *tu/vous* in Pau, but neither of these factors was significant, despite trending in predictable directions (SD more favored with *tu* and with younger speakers).

An analysis of both groups of Near-NSs combined may help to explain these findings. As with *ne*-retention, Near-NSs at both sites were combined in an Rbrul analysis. Table 5-28 provides the significant factor groups for this grouping, consisting of 19 speakers across 38 conversations.

Table 5-28. Significant factor groups in Near-NSs for Pau and Lille combined (subject doubling)

Factor Group	Factor	Log odds	Factor weight	% SD	N
Clitic context[107]	neuter	1.492	.816	77.5	183/236
	ambiguous	-0.212	.447	42.4	75/177
	personal	-1.280	.218	35.6	233/655
Subject type[108]	strong pronoun	2.308	.910	91.9	57/62
	verb/PP	0.161	.540	92.6	25/27
	proper noun	-0.738	.324	38.9	63/162
	indefinite/quantifier	-0.758	.319	38.2	34/89
	common noun	-0.973	.274	42.9	312/728
Negation	*ne*-deletion	1.240	.776	71.0	49/69
	affirmative	0.285	.571	46.1	435/943
	ne-retention	-1.525	.179	12.5	7/56
Clause type	matrix	1.210	.770	51.3	406/791
	subordinate	0.570	.639	33.8	80/237
	si	-0.765	.318	18.2	4/22
	relative	-1.015	.266	5.6	1/18
Relative clause	relative clause	0.683	.664	73.0	54/74
	no relative clause	-0.683	.336	44.0	437/994
Subject animacy[109]	animate	0.779	.685	42.1	236/561
	inanimate & material	0.167	.542	44.1	15/34
	inanimate & immaterial	-0.207	.448	54.2	195/360
	place	-0.739	.323	39.8	45/113
Verb type	copula	0.679	.664	57.1	281/492
	transitive	0.093	.523	38.5	84/218
	intransitive	0.082	.520	35.8	97/271
	modal	-0.125	.469	38.7	24/62
	passive	-0.729	.325	20.0	5/25
Subject definiteness[110]	definite	0.527	.629	46.8	422/901
	verb (not applicable)	0.285	.571	92.0	23/25
	indefinite	0.101	.525	47.5	28/59
	quantified	-0.913	.286	21.7	18/83
TOTAL				46.0%	491/1068

Input probability = 0.365; Log likelihood = -530.798

Non-significant factor groups (cf. Table G-14): Preverbal clitics, Other preverbal material, Subject specificity, Conversation portion, *Tu/vous*, Interlocutor L1/identity, Site

[107] The VIF for this factor group was above 2.5.

[108] The VIF for this factor group was above 10.

[109] The VIF for this factor group was above 2.5.

[110] The VIF for this factor group was above 10.

For Near-NSs observed as a single group, there is a re-organization of several factor groups identified as significant for each site. From Table 5-27, we can compare the significant factor groups that remain when combining all Near-NSs (Table 5-29).

Table 5-29. Significant factor groups for Near-NSs in both sites and overall (subject doubling)

Near-NSs in Lille	**Near-NSs in Pau**	**Near-NSs overall**
Clitic context	Clitic context	Clitic context
Subject type	Subject type	Subject type
Clause type	Negation	Negation
Subject animacy	Relative clause	Clause type
Verb type	Subject animacy	Relative clause
	Clause type	Subject animacy
	Subject specificity	Verb type
	Verb type	Subject definiteness

Unsurprisingly, Clitic context and Subject type remain the two most significant factors. For Clitic context, we also see a more robust difference between ambiguous and personal contexts, with more inhibition of SD in personal contexts. All of the rankings in Lille remain in the same order when combined with Pau, whereas the combined rankings elevate Clause type compared with Pau alone. Furthermore, while Subject animacy remains significant, Subject specificity drops from the rankings, and Subject definiteness, with more overall tokens, now becomes significant, albeit ranked the lowest. Relative clauses and negation are the pre-verbal elements that significantly influence SD when both groups are combined; there is little impact of other intervening material.

No extralinguistic/sociostylistic factors were selected as significant. It is somewhat surprising that *Tu/vous* was not significant, since combining the sites increases the differences found in Near-NSs in Pau. The influence of this factor can nevertheless still play a role in SD rates, whether it is the act of selecting a formal versus informal address pronoun or, more generally, the adoption of a more formal versus informal style that conditions this sociolinguistic variable.

Importantly, the testing site (Pau versus Lille) was not selected as a significant factor, suggesting that, despite differences in the SD rates for each site (42% versus 53%), linguistic factors essentially account for the variation found in this overall grouping.

5.5.5.1 Results: Interlocutors with Near-NSs

Table 5-30 provides the significant factor groups for the two native speaker interlocutors in Lille across 9 Near-NS/NS conversations. Due to low token counts, several factors were regrouped after an initial Rbrul run. In the Pre-verbal clitics group, reflexives (5 tokens; no SD) were grouped with object clitics; in the Other pre-verbal material group, parentheticals and adverbs were grouped together (14 tokens; all SD), and hesitations (6 tokens; 3 SD) were grouped with tokens containing no pre-verbal material; in the Clause type group, relatives (3 tokens; no SD) and *si* clauses (3 tokens; 2 SD) were grouped with subordinate clauses.

Table 5-30. Significant factor groups for NS interlocutors in Lille (subject doubling)

Factor Group	Factor	Log odds	Factor weight	% SD	N
Pre-verbal clitics	no clitic	1.831	.862	59.5	150/252
	object/reflexive clitic	-1.831	.138	6.2	1/16
Clitic context	neuter	1.003	.732	82.1	55/67
	ambiguous	-0.374	.407	50.8	30/59
	personal	-0.628	.348	46.5	66/142
Other pre-verbal material	parenthetical/adverb	11.439	> .999	100.0	14/14
	prepositional phrase	-5.034	.006	75.0	15/20
	hesitations/none	-6.405	.002	52.1	122/234
Clause type	matrix	0.517	.626	66.3	120/181
	subordinate/relative/*si*	-0.517	.374	35.6	31/87
Subject type	verb/PP	13.178	> .999	100.0	10/10
	strong pronoun	-1.264	.220	85.7	12/14
	common noun	-3.909	.020	52.5	106/202
	proper noun	-3.919	.019	56.7	17/30
	indefinite/quantifier	-4.086	0.17	50.0	6/12
Negation	*ne*-deletion	7.092	.999	72.2	13/18
	affirmative	5.469	.996	56.6	138/244
	ne-retention	-12.562	< .001	0.0	0/6
TOTAL				56.3%	151/268

Input probability = 0.952; Log likelihood = -131.039
Non-significant factor groups (cf. Table G-15): Relative clause, Verb type, Subject definiteness, Subject specificity, Subject animacy, Conversation portion

Regarding overall ranking of significant factor groups, compared with Near-NSs and SA interlocutors, NSs in Lille appear to demonstrate a stronger sensitivity to pre-verbal material across the board, with Pre-verbal clitics as the most significant factor group (albeit with low token counts) and other pre-verbal material influencing SD variation more strongly than Subject type. Otherwise, the results for this NS group (with speaker CaF accounting for most of the tokens) broadly align with previous studies (though note that SD frequency with passive verbs (50%) is higher than in any other study). Note, furthermore, the slight inhibiting effect of Clitic context in personal pronoun contexts compared with ambiguous contexts, as seen in the Near-NS groups. As for NSs in Pau, Table 5-31 details the significant factor groups for the three native speaker (bilingual) interlocutors across 20 Near-NS/NS conversations.

Table 5-31. Significant factor groups for NS interlocutors in Pau (subject doubling)

Factor Group	Factor	Log odds	Factor weight	% SD	N
Clitic context[111]	neuter	1.646	.838	87.8	158/180
	ambiguous	-0.562	.363	50.8	62/122
	personal	-1.084	.253	57.7	319/553
Clause type	matrix	1.163	.762	69.6	455/654
	subordinate	0.008	.502	44.4	75/169
	relative	-0.130	.468	44.4	4/9
	si	-1.041	.261	21.7	5/23
Subject type	verb/PP	10.678	> .999	100.0	16/16
	strong pronoun	-0.674	.338	87.9	29/33
	proper noun	-2.717	.062	60.2	80/133
	common noun	-3.012	.047	62.2	399/641
	indefinite/quantifier	-4.275	.014	46.9	15/32
Negation	*ne*-deletion	1.126	.755	73.6	53/72
	affirmative	0.811	.692	62.9	484/769
	ne-retention	-1.936	.126	14.3	2/14
Subject animacy[112]	inanimate & material	0.571	.637	78.9	15/19
	animate	0.455	.612	61.5	312/507
	inanimate & immaterial	0.056	.514	69.7	184/264
	place	-1.073	.255	43.1	28/65
Relative clause	relative clause	0.911	.713	86.5	32/37
	no relative clause	-0.911	.287	62.0	507/818
Other pre-verbal material	parenthetical	1.829	.862	94.7	18/19
	hesitations	0.945	.720	76.2	16/21
	prepositional phrase	-0.186	.454	67.2	39/58
	none	-0.457	.388	62.0	448/723
	multiple elements	-0.577	.360	50.0	7/14
	adverbs	-1.553	.175	55.0	11/20
Interlocutor identity	English identity	0.207	.552	64.7	317/490
	French identity	-0.207	.448	60.8	222/365
TOTAL				63.0%	539/855

Input probability = 0.965; Log likelihood = -419.441
Non-significant factor groups (cf. Table G-16): Verb type, Subject definiteness, Subject specific, Preverbal clitics, Conversation portion, *Tu/vous*

As with the other speaker groups in the current study, for bilinguals in Pau, Clitic context and Subject type are high-ranking factor groups, though Clause type appears higher for this speaker

[111] The VIF for this factor group was above 2.5.
[112] The VIF for this factor group was above 2.5.

group than for any others. For Clitic context, these bilinguals also show a slightly more inhibiting effect for ambiguous contexts than for personal contexts, reversing the slight trend seen in all other near-native and NS groups. We also see a rare example of a significant extralinguistic factor group for these speakers, Interlocutor identity. Though a *chi*-square test determined that there were no significant differences in SD rates when these NSs adopted an English versus French identity, this factor group appears as significant (though the least significant of all such factor groups) in the variationist analysis. The fact that interlocutor identity did not appear as significant for the Near-NSs in Pau (41.8% SD with interlocutor in English identity versus 42.0% with interlocutor in French identity) suggests that this apparent difference in SD usage by the bilinguals across both identities did not influence SD rates in Near-NSs.

When comparing each group of NSs in the current study (the NSs in Lille and the bilinguals in Pau), we see broadly similar trends when compared with previous analyses of SD in NSs. For Clause type, the observation by Nagy et al. (2003) that *si* clauses favor SD the least is supported by the current analysis for NSs as well. Subject type is among the most significant factor groups, and the ordering of factors in both sites is as predicted on the continuum seen in Nadasdi (1995b), with the exception in the current study that factor weights of proper nouns and common nouns are barely distinguishable. Negation behaves in a similar way as well; though it is ranked lower in Lille, the same trend seen in all other studies obtains: *ne*-retention disfavors SD while *ne*-deletion favors it. One difference from previous studies (e.g., Nadasdi, 1995b) concerns pre-verbal material: in Pau, intervening adverbs are among the least favorable to SD and rank lower than hesitations, whereas in Lille and in Nadasdi's study, the opposite ordering is obtained. Again, low token counts for both factors may explain this discrepancy.

Most of the other subject factors (definiteness, specificity, animacy) reveal few discrepancies from previous studies on NSs: definiteness favors SD while indefinites and quantifiers disfavor SD; specific subjects slightly favor SD while non-specific and generalized subjects slightly disfavor SD. However, with the exception of subject animacy in Pau, the number of tokens that disfavor SD is not enough to make these distinctions significant in each of these NS groups.

As with Near-NSs, an Rbrul analysis was run for NSs at both sites together. Table 5-32 provides the significant factor groups, involving five speakers and 29 conversations (20 in Pau, 9 in Lille).

Table 5-32. Significant factor groups for NS interlocutors in Lille and Pau (subject doubling)

Factor Group	Factor	Log odds	Factor weight	% SD	N
Clitic context[113]	neuter	1.309	.787	86.2	213/247
	ambiguous	-0.551	.366	50.8	92/181
	personal	-0.758	.319	55.4	385/695
Clause type	matrix	1.059	.742	68.9	575/835
	subordinate	-0.006	.499	42.1	104/247
	si	-0.503	.377	24.1	7/29
	relative	-0.550	.366	33.3	4/12
Negation	*ne*-deletion	1.344	.793	73.3	66/90
	affirmative	0.742	.677	61.4	622/1013
	ne-retention	-2.086	.110	10.0	2/20
Subject type	verb/PP	11.506	> .999	100.0	26/26
	strong pronoun	-1.258	.221	87.2	41/47
	proper noun	-3.109	.043	59.5	97/163
	common noun	-3.298	.036	59.9	505/843
	indefinite/quantifier	-3.841	.021	47.7	21/44
Other pre-verbal material	parenthetical	2.138	.895	96.4	27/28
	hesitations	0.261	.565	70.4	19/27
	prepositional phrase	0.132	.533	70.1	54/77
	none	-0.590	.357	59.6	567/951
	multiple elements	-0.883	.293	64.0	16/25
	adverbs	-1.058	.258	46.7	7/15
Relative clause	relative clause	0.760	.681	85.5	47/55
	no relative clause	-0.760	.319	60.2	643/1068
Pre-verbal clitics	none	0.693	.667	63.2	658/1041
	object clitic	-0.095	.476	40.7	24/59
	reflexive clitic	-0.599	.355	34.8	8/23
Subject specificity	specific	0.395	.598	66.3	579/873
	non-specific	-0.048	.488	43.9	47/107
	generalizing	-0.347	.414	44.8	64/143
Subject animacy[114]	inanimate & material	0.550	.634	71.0	22/31
	animate	0.119	.530	58.6	365/623
	inanimate & immaterial	0.014	.504	67.6	248/367
	place	-0.684	.335	53.9	55/102
Gender	male (*n*=3)	0.578	.641	70.8	433/612
	female (*n*=2)	-0.578	.359	50.3	257/511
TOTAL				61.4%	690/1123

Input probability = 0.946; Log likelihood = -552.956

Non-significant factor groups (cf. Table G-17): Verb type, Subject definiteness, Conversation portion, *Tu/vous*, Site

[113] The VIF for this factor group was above 2.5.

[114] The VIF for this factor group was above 2.5.

What is particularly striking in this analysis of all NSs is the number of significant factor groups (10), higher than for any other speaker grouping. As with the combination of all Near-NSs, the addition of more tokens allows for previously non-significant factor groups to appear in these rankings. For ease of comparison, Table 5-33 shows only the significant factor group rankings in order for NSs.

Table 5-33. Significant factor groups for NSs (subject doubling)

NSs in Lille	**NSs in Pau**	**NSs overall**
Pre-verbal clitics	Clitic context	Clitic context
Clitic context	Clause type	Clause type
Other pre-verbal material	Subject type	Negation
Clause type	Negation	Subject type
Subject type	Subject animacy	Other pre-verbal material
Negation	Relative clause	Relative clause
	Other pre-verbal material	Pre-verbal clitics
	Interlocutor identity	Subject specificity
		Subject animacy
		Gender

As with Near-NSs, all of the factor groups in Lille appear in Pau, but the overall ordering more closely reflects that obtained in Pau, which is expected given the larger token numbers in Pau. However, the overall rankings emphasize one pre-verbal element (negation) more strongly with the two sites combined, while de-emphasizing other pre-verbal elements. Also, with five total speakers, it is possible to add Gender as a factor group, which was the least significant factor group but nevertheless reflects the fact that the speakers with the two highest SD rates were both female (CaF and Ch), consistent with Auger and Villeneuve's (2010) findings concerning gender and age. As with Near-NSs, the testing site (Pau versus Lille) was not selected as a significant factor, with a comparatively smaller difference in the SD rates for each site (63% versus 56%).

Other trends also appear more robust when the data from each site are combined. While Clause type generally follows the same order as in previous studies, the NSs in the current study favor SD more in non-matrix clauses overall (40%) compared with previous studies (28% in Auger & Villeneuve (2010) and 10% in Zahler (2014)). One surprising result is that in the combined data, though quantified subjects disfavor SD as in previous studies, and the Subject definiteness group was not selected as significant, definite and indefinite subjects essentially pattern identically, with indefinites neither favoring nor disfavoring SD. Though low token counts may account for this result, it is possible that, diachronically, indefinites are becoming more frequently doubled.

5.5.5.2 Results: Interlocutors (near-native) in Lille

The last speaker group concerns the two near-native interlocutors in Lille. Table 5-34 provides the significant factor groups for these speakers. These results involve 9 Near-NS/near-native conversations. Due to low token counts, the following factors were regrouped after an initial Rbrul run: in the Subject type category, strong pronouns (5 tokens; all SD) and infinitival subjects (10 tokens; all SD) were grouped together; in the Pre-verbal material category, one adverb token (with SD) was grouped with parentheticals (4 tokens; all SD).

Table 5-34. Significant factor groups for near-native interlocutors in Lille (subject doubling)

Factor Group	Factor	Log odds	Factor weight	% SD	N
Clitic context	neuter	1.808	.859	93.8	45/48
	ambiguous	-0.586	.358	54.1	20/37
	personal	-1.222	.228	40.0	52/130
Subject type	strong pronoun/verb/PP	12.275	> .999	100.0	15/15
	proper noun	-3.259	.037	69.6	16/23
	common noun	-3.980	.018	47.3	79/167
	indefinite/quantifier	-5.036	.006	70.0	7/10
Other pre-verbal material	parenthetical/adverb	13.800	> .999	100.0	6/6
	prepositional phrase	-2.373	.085	76.9	10/13
	multiple elements	-3.174	.040	66.7	4/6
	none	-3.911	.020	53.2	91/171
	hesitation	-4.343	.013	31.6	6/19
Relative clause	relative clause	0.731	.675	85.0	17/20
	no relative clause	-0.731	.325	51.3	100/195
TOTAL				54.4%	117/215

Input probability = 1.000; Log likelihood = -105.801
Non-significant factor groups (cf. Table G-18): Clause type, Verb type, Subject definiteness, Subject specificity, Subject animacy, Preverbal clitics, Negation, Conversation portion

With a small number of tokens compared with other Rbrul runs, it is unsurprising that fewer factor groups are found to be significant for these interlocutors. Broadly, the same trends as those seen in the other speaker groups seem to obtain in these speakers, and Clitic context and Subject type again dominate the significant factor groups. For this speaker group, however, Clitic context yields the largest difference in ambiguous versus personal contexts, with the strongest inhibition of SD in personal contexts. For Subject type, SD is favored with proper nouns over common nouns at a rate that is closer to observations from previous studies compared with the other speaker groups in the current study, where the differences between these two subject types were more marginal. Though low token counts preclude definitive conclusions, one exception to the broadly similar trends observed in this group is for Subject specificity, in which [-specific] nouns did not inhibit SD at anywhere near the rates obtained in previous studies. A plausible explanation for these results may be found in the conversational context for speaker JeE, who

produced the majority of all generalizing tokens for this speaker group within a single conversation in which stereotypes of French people was the predominant topic, and in which SD was produced with nine of 13 generalizing tokens.

5.5.6 Summary of variationist analyses on subject doubling

Across all speaker groups from the current study, several observations can be made from these variationist analyses. Properties of the subject, such as Clitic context and Subject type, consistently appear as the most significant factor groups for highly proficient speakers. Moreover, the linguistic constraints in Near-NSs largely pattern like NS constraints, though non-copula verbs inhibit SD to a greater extent in Near-NSs than in NSs, leading Verb type to be a significant factor group for the former speaker group but not the latter. In SA learners, properties of the verb are the most significant factors, with copulas (such as lexicalized *c'est*) and matrix clauses strongly favoring SD, and non-copulas and non-matrix clauses inhibiting SD. Elsewhere, pre-verbal material that does not consist of subject relative clauses, object or negation clitics also appears to more strongly condition SD in NSs, with a range of factors favoring and slightly disfavoring SD (ranging from 46% to 96%), whereas Near-NSs show less influence from these factors (ranging from 40% to 63%). Indefinites and quantified subjects also trend toward higher SD rates for both Near-NS and NS groups compared with previous studies.

Crucially, interlocutor language background is a significant factor group influencing the SD variable for SA learners (i.e., higher SD frequency indicating more informal usage with the NS and near-native interlocutor compared with another SA learner), in the same direction as the effect observed in the variationist analysis for the *ne*-retention variable in Chapter 4. In Near-NS groups, the interlocutor language background was not a significant factor group conditioning SD variation, despite some differences in SD frequency across interlocutor type observed in the Lille

Near-NSs. Other social factors appear to influence variation to a minor extent in some groups, but not to the same extent as the factors outlined above.

5.6 Hypotheses revisited: subject doubling

I return here to the research questions and hypotheses outlined in Chapter 3, with respect to subject doubling. Recall that the first research question probes the sociolinguistic ability of L2 French learners at various proficiency levels, with the hypothesis that intermediate/advanced learners will demonstrate greater adherence to standard norms than Near-NSs with regard to sociolinguistic variables, and that near-natives may approach NS patterns but still demonstrate non-targetlike deviation from these patterns. For subject doubling, hypothesis #1 was confirmed. SA learners had significantly lower overall SD rates than their near-native and native interlocutors. For the Near-NSs, SD rates vary by site: Near-NSs in Pau had significantly lower SD compared with NS interlocutors, whereas in Lille the Near-NSs had slightly lower SD rates and a non-significant difference compared to their NS interlocutors. The variationist analyses also confirmed strong linguistic constraint similarities between Near-NSs and NSs, but less so between SA learners and their more proficient interlocutors.

Recall that the second research question asks how much the interlocutor plays a role in influencing the production of sociolinguistic variables such as SD, with the hypothesis that lower-level learners will be expected to show convergence toward their interlocutor in usage of sociolinguistic variables, while Near-NSs are expected to be less influenced by interlocutor effects, though they may only be able to exhibit targetlike sociolinguistic behavior in interaction with NSs (thus possibly revealing a secondary interlocutor effect). Hypothesis #2 was confirmed for SA learners with subject doubling. These learners showed evidence of convergence, with significantly higher SD with near-native and native speakers than with other SA learners, and, as with *ne*-

retention, the Rbrul analysis identified interlocutor status as the lone significant extralinguistic factor. As for Near-NSs, the two testing sites show differences with respect to interlocutor effects on SD: Near-NSs in Pau had essentially identical SD rates when speaking to the bilinguals adopting either identity, whereas Near-NSs in Lille had significantly higher SD rates in conversation with NSs compared with other near-native interlocutors. It appears, then, that for the latter site, some combination of two factors are in play: nativelike SD rates are facilitated in Near-NSs through conversation with NSs (even when this is a bilingual assuming an L1 English identity), or nativelike SD rates are inhibited by speaking to a near-native with clear non-targetlike phonological and syntactic patterns. These discussions will be revisited in the final chapter.

Chapter 6: *Ne*-retention/subject doubling interactions and conclusions

The main objective for the current study was to examine the effect of interlocutor language background on the use of sociolinguistic variation in adult learners of French. In this final chapter, I begin with a brief summary of the study and its principal conclusions (section 6.1). I then examine in more detail the interaction between the two sociolinguistic variables examined in this study (sections 6.2 and 6.3). Finally, I discuss the findings from the current study in light of their pedagogical and methodological implications for second language acquisition (sections 6.4 and 6.5), and I conclude with final observations and avenues for future research (section 6.6).

6.1 L2 speakers of French and the interlocutor effect: A summary

Having identified the interlocutor language background as an inconsistently monitored feature of previous studies of L2 sociolinguistic performance, I posed research questions concerning the nature of L2 sociolinguistic variation and the potential effect of the interlocutor language background. I manipulated this interlocutor dimension in specific ways by recruiting native, bilingual, and near-native speakers of French to serve as conversation partners for learners of French. After verifying learner proficiency levels and learner motivations through written tasks, I collected a new corpus involving dyadic conversational data obtained from learners at a commonly studied level of proficiency (a low-advanced group with study-abroad experience) and learners at a less-commonly studied level (highly proficient, near-native speakers), in order to determine if any such interlocutor effect was detectable across multiple proficiency levels. An examination of the corpus determined that the conversations consisted of an informal style in which sociolinguistically-conditioned informal variants were anticipated to occur, and my quantitative, variationist analyses focused on two sociolinguistic variables in French: a more

commonly studied variable in L2 variation (*ne*-retention) and a less-commonly studied variable (subject doubling).

Here, I revisit the significant findings regarding the effect of the interlocutor language background on SA learners for the two sociolinguistic variables examined in the current study. For *ne*-retention, Table 4-7 is reproduced as Table 6-1 here, and for SD, Table 5-8 is reproduced as Table 6-2.

Table 6-1. SA learners' use of ne-retention by interlocutor type

Interlocutor type	***ne* / total negation tokens**	**% *ne*-retention**
L1 French: NS	90/162	55.6
L2 French: near-native	101/169	59.8
L2 French: SA learner	133/163	81.6

Table 6-2. SA learners' use of SD by interlocutor type

Interlocutor type	**SD / Total NPs**	**% SD**
L1 French: NS	34/119	28.6
L2 French: near-native	53/172	30.8
L2 French: SA learner	11/125	8.8

SA learners' use of the two variables is strikingly similar: informal variants (absence of *ne* and presence of SD) are favored when in conversation with a native speaker and with a highly proficient speaker, while formal variants are favored when in conversation with another SA learner. Based on these results, I concluded that SA learners considered the NS and near-native as model speakers of French, since these speakers were integrated into the Francophone community and thus represented the type of speech that motivated learners may wish to emulate. For the SA learners for whom variation in *ne*-retention and subject doubling is part of their sociolinguistic repertoire, there is evidence of an interlocutor effect that, along with other social and contextual factors (informal conversation, similarities in age, education, and motivation), presents as close to

an ideal situation for measuring nativelike use of sociolinguistic variables as an individual measure of recorded spontaneous oral production can obtain. This interlocutor effect appears to prime learners at this proficiency level for more nativelike use of sociolinguistic variation when speaking with NSs and near-natives. For many SA learners (though not all), this effect results in convergence with highly proficient speakers regarding multiple sociolinguistic variables. Furthermore, the SA learners' strikingly similar usage of the informal variant in conversation with the native *and* near-native interlocutor indicate that high proficiency in French, whether native or not, may be sufficient for learners to produce higher frequencies of colloquial variants compared to SA learner dyads. Indeed, in the absence of a native or near-native model speaker, learners appear to set an alternative norm for sociolinguistic behavior and accommodate to their fellow SA interlocutors' speaking style. These factors may be mutually reinforced (consistent with the Interactive Alignment model), resulting in a stronger adherence to more formal variants and a more classroom-like speaking style.

While the debriefings revealed that learners generally did not have difficulty expressing themselves in conversation with other learners, this relative ease in SA learner dyads did not necessarily translate to more nativelike sociolinguistic variation. For example, SA learner 5S reported feeling more pressure when conversing with native speakers in general, yet her use of both *ne*-retention and SD was more nativelike with the NS than with an SA learner. Furthermore, in the debriefing, five of the eight SA learners cited the NS as the most difficult conversation, due mainly to the NS' speech rate and vocabulary, though *ne*-retention and SD rates for this subset were more nativelike with the NS than with the other two interlocutors.

As discussed in section 4.5.7.1, Dewaele's (2004) categorization of "L2 learners" in interaction with fellow learners compared to "legitimate L2 *users*" (italics mine) in interaction

with NSs implies a distancing effect on the learner's relationship with the target language. It is unlikely that the SA learners, for example, would be aware of such labels, but these distinctions may surface in their motivations for selecting which language to use in a given interaction (a conscious choice) and which style of language to use in a given interaction (not necessarily a conscious choice). The somewhat "artificial" context of speaking in an L2 when both learners share the same language may cause one or both learners to exert more conscious selection of certain forms across different linguistic domains. This selection may be due to a desire, for example, to speak more standard French to make sure that a less proficient interlocutor can understand, just as much as it can be due to the speaker feeling self-conscious about appropriating (at least temporarily) an alternate identity (as a *user* of French), creating distance from the initial shared identity of both learners as L1-English speaking Americans. On the other hand, Sax (personal communication, July 17, 2020) posits that there may be "competition" among groups of learners to determine who speaks the best or most nativelike, which might result in learners trying to speak more nativelike with fellow learners. However, some learners at a proficiency level similar to these SA learners (and indeed, even high-proficiency learners) may still consider the standard form to be the "best" form of the language when under evaluation in a recorded setting, and they may be concerned by stylistically inappropriate use of informal forms when the potential payoff of using such forms (such as cultural convergence with a native speaker) is reduced in SA learner dyads.

The near-categorical maintenance of formal variants (retention of *ne* and lack of SD) in certain learners, such as speakers 6S and 7S, likely indicates a lack of conscious attention to variation contexts, even when exposed to variation by their interlocutor in a relatively "safe" environment favoring informal communication (that is, learners were encouraged to have a casual

conversation, and the interlocutors adopted an informal tone; learners did not necessarily have to determine the informality of the communicative context on their own). Though the debriefings revealed that none of the learners claimed to actively monitor their use of *ne*, they all claimed awareness of its use in formal contexts. Furthermore, two learners (2S and 6S) mentioned being aware that their use of *ne* had decreased since their arrival in France, though, paradoxically, 6S produced essentially categorical retention in his conversations.

Global awareness of stylistic variation also appears to compete with specific linguistic contexts. Lexicalization was one of the most robust predictors of *ne*-deletion rates in learners, as the data in Chapter 4 indicate. Lexicalization involves the use of more formulaic routines; while each utterance of a lexicalized form (e.g., *c'est pas*) has the potential to be analyzed as an environment for variation (e.g., selecting *ce n'est pas* instead), it is unlikely that this is a conscious selection for each such utterance, especially after the formality of the conversation has been established, and especially when lack of conscious selection can have cognitive advantages for learners. Conversely, learners were not on "autopilot" for all lexicalized forms, as evidenced by the 74% *ne*-retention for these forms in the variationist analysis. In a similar way, subject doubling with *c'est* requires comparatively few processing demands and is indicative of some level of formulae even in learners at this level, but it is far from the near-categorical use observed in native speaker surveys (cf. Barnes, 1985; Coveney, 2003). Elsewhere, it is likely that development of proficiency, resulting in reduced cognitive load, may lead to increased use of informal variants; where such variants introduce competition with the learner's previous default (viz., formal) forms, this may heighten the awareness of a conscious selection of either form in certain learners now more attuned to the style expected of a certain social interaction. This competition then may be subject to increased sensitivity to interlocutor effects, given the learner's increased awareness of

her own variation and awareness of how it may align with or differ from the style used by her interlocutor.

As for near-native speakers, data on *ne*-retention and SD in the current study indicate that such speakers' sociolinguistic behavior does not seem to be as influenced by the interlocutor effect of language background. Though the results indicate that there may be some effect for SD, broadly it appears that nativelike sociolinguistic performance on these variables is not necessarily inhibited by speaking to another non-native at a similar proficiency level; that is, a native speaker does not necessarily need to be in the sphere of conversation in order for Near-NSs to exhibit nativelike sociolinguistic behavior, based on the non-significant differences across interlocutor types for *ne*-retention and, in the case of Near-NSs in Pau, for SD. Furthermore, the adoption of an L1 English or L1 French identity by bilinguals had no significant effect on either variable, suggesting that any interlocutor effect at this level requires reinforcement of the interlocutor's native language status by the interlocutor's native or non-native linguistic production. Highly proficient speakers may have already reached a certain comfort level in the use of sociolinguistic variants, and at sufficiently advanced proficiency, sensitivity to interlocutor characteristics (or, more narrowly, to interlocutor language background) may not be reflected in this behavior when in an informal speaking context. Furthermore, high proficiency may allow the learner to more easily identify the appropriate stylistic variation for this type of informal interaction, but her output may be conditioned by a potentially stable baseline level of use of sociolinguistic variables that is not necessarily the same as that used by native speakers, for all variables, in such an interaction.

In sum, as the results indicate in Chapters 4 and 5, SA learners show evidence of sensitivity to the interlocutor language background effect with regard to two individual sociolinguistic variables, while Near-NSs do not necessarily demonstrate such sensitivity. While I have touched

upon the co-occurrence of these two variables in the respective chapters where relevant, more insights into these speakers' sociolinguistic profiles can be revealed upon a closer examination of the co-occurrence of *ne*-retention and subject doubling. The following two sections will elaborate on the behavior of the current study's three main groups of speakers in these contexts and the acquisitional challenges faced by learners therein.

6.2 *Ne*-retention use in SD contexts

As mentioned in the introduction to Chapter 5, *ne*-retention and SD occur in the same morphosyntactic "neighborhood." Since both variables can appear in the same utterance (and the doubling subject clitic and *ne* can occur as consecutive morphemes), nativelike distribution of these two variables in possible co-occurrence contexts can be an acquisitional challenge for learners, from a sociolinguistic perspective. To address the question concerning the extent to which learners can master the complexities of variable sociolinguistic contexts, it would be useful to analyze the distribution of *ne*-retention and SD in negative clauses that contain a lexical or strong pronoun subject (e.g., *Ma mère ne parle pas français* 'My mother does not speak French') in both native and non-native speakers of French.

By way of background, Villeneuve and Auger (2013) is one of the few previous studies to have examined the interaction of both of these variables in an informal context with speakers of French (whom the authors examine in comparison with bilingual speakers of French and Picard). This study was inspired by claims in Massot (2010) and Zribi-Hertz (2011) that *ne*-retention and SD are incompatible. Since bipartite negation is much more common in Picard than in informal French, *ne*-retention frequently co-occurs with SD in speakers of Picard. However, Villeneuve and Auger find that the same bilingual speakers retain *ne* much more rarely in SD contexts in spoken French, at similar rates to a monolingual French control group, and the authors advance an

argument that the different distribution patterns of SD and *ne*-retention within these two Gallo-Romance languages belong to different grammars rather than to the same grammar governing both varieties. Despite differences in methodological decisions concerning the contexts for analyzing both variables compared with the current study (for example, the authors excluded collocations such as *il y a*, *c'est*, and *il faut* from their *ne*-retention results), and despite different demographic characteristics in their speakers (average age of monolinguals = ~50), we can compare the interaction of these variables in the current study with such an analysis. For the four monolinguals sampled in Villeneuve and Auger's study, average *ne*-retention was 31.5% (considerably higher than the average of any Near-NS or NS group in the current study, though likely due in part to the exclusion of collocations); average SD was 42.3%, slightly lower than overall NS rates in the current study but not significantly lower than the NS rates in Lille ($\chi^2(1) = 3.17$; $p = .075$).[115]

The percentage of each type of distribution of these two variables in Villeneuve and Auger also suggested no difference between monolinguals and bilinguals; when examining tokens in which possible SD occurs in a negative clause (with possible *ne*-retention/deletion), the distribution in Table 6-3 obtains:

Table 6-3. Distribution of SD + negation contexts in Villeneuve & Auger (2013)

SD & negation patterns	**Monolinguals** (total tokens = 20)	**Bilinguals** (total tokens = 32)	**Example**
1. no SD; *ne*-retention	11 (55%)	20 (63%)	*Ma mère ne parle pas français.* 'My mother does not speak French.'
2. SD; *ne-deletion*	4 (20%)	7 (22%)	*Ma mère elle parle pas français.* 'My mother (she) doesn't speak French.'
3. no SD; *ne-deletion*	4 (20%)	4 (13%)	*Ma mère parle pas français.* 'My mother doesn't speak French.'
4. SD; *ne*-retention	1 (5%)	1 (3%)	*Ma mère elle ne parle pas français.* 'My mother (she) does not speak French.'

[115] Note that, in addition to lexical subjects, Villeneuve and Auger (2013) included all 3rd person strong pronoun subjects in the spoken French data (*lui, elle, eux, elles*), whereas the current study excluded *elle(s)*.

Despite the small token counts for these contexts, the distribution of the two variables is similar for both speaker groups. Villeneuve and Auger (2013) conclude that SD and *ne*-retention are not incompatible; rather, this combination is merely disfavored, as it involves variants with different formality levels, but their variationist approach predicts rare tokens of such co-occurrences.

Other studies have also examined SD and *ne*-retention in similar contexts, though small token counts also make definitive conclusions more difficult. Meisner (2016) goes further than Villeneuve and Auger by suggesting that there is an incompatibility with SD and *ne*-retention (cf. Massot, 2010), as her corpus contains 22 tokens of SD in negation contexts, but with categorical *ne*-deletion. Stark (2012) arrives at the same conclusion for her corpus of text messages; in negation contexts, all 8 tokens of SD produce *ne*-deletion. Elsewhere, Hansen and Malderez (2004: 21), in their study on *ne*-retention, report data that can be converted to a similar format as in Table 6-3. For 76 tokens of possible doubling with lexical subjects in negations contexts, the percentages break down as follows: 1) 41% (31/76); 2) 25% (19/76); 3) 32% (24/76); 4) 3% (2/76). Hansen and Malderez also comment on the influence of SD from the perspective of negation contexts consisting of subject clitics alone (non-doubled) compared with negations contexts consisting of SD (lexical NPs + subject clitic anaphors); the former contexts show 5.8% retention overall (for 1204 tokens) compared with 9.5% retention in the latter (for 21 tokens), suggesting that the presence of a dislocated subject NP only has a minimal effect on *ne*-retention with subject clitics.

In the current study, the variationist analysis of SD identified negation contexts as a significant factor group for only a subset of the speaker groups (Near-NSs in Pau; Near-NSs overall; NSs for each site and overall). However, all of the speaker groups trend in the same direction with regard to SD usage in negation contexts: more SD in clauses in which *ne* is omitted than in clauses where it is overt (see Table 6-4).

Table 6-4. SD usage in negation contexts

Speaker group	SD tokens / total variable contexts; % SD		SD tokens; % SD in negation contexts with *ne*-retention		SD tokens; % SD in negation contexts with *ne*-deletion	
SA learners	98/416	23.6	0/18	0.0	8/11	72.7
SA interlocutors	93/141	66.0	0/1	0.0	8/9	88.9
Near-NS: Pau	287/685	41.9	4/45	8.9	28/38	73.7
Near-NS: Lille	204/383	53.3	3/11	27.3	21/31	67.7
NS: Pau	539/855	63.0	2/14	14.3	53/72	73.6
NS: Lille	151/268	56.3	0/6	0.0	13/18	72.2
NNS: Lille[116]	117/215	54.4	3/7	42.9	5/7	71.4
Total	1489/2963	50.3	12/102	11.8	136/186	73.1

First, the appearance of negation contexts does little to affect overall SD rates (50.3% overall versus 51.4% in negation contexts), since negation contexts pull in both directions on SD, and the higher number of tokens in *ne*-deletion contexts is counteracted by the SD percentage in this context being closer to the overall percentage. Importantly for the acquisition process, all groups of learners pattern like natives: lower SD in *ne*-retention contexts and higher SD in *ne*-deletion contexts. This suggests that learners have correctly identified that SD and *ne*-deletion broadly go hand-in-hand in informal styles. What is more striking, however, is the consistent SD rate in *ne*-deletion contexts, with a range of only 21% across all groups (67.7% to 88.9%). Despite the small token numbers, even the SA learners pattern exactly like native speakers here: in the few contexts of SD in SA learners, they also tend to drop *ne*. However, what distinguishes SA learners from NSs is that, in these 11 SD tokens for SA learners, there is a complementary distribution of doubling with *c'est* (with seven tokens of *c'est pas* and one *c'est rien*) and non-doubling with other verbs (three tokens, comprising the verbs *parle, sont,* and *fait*), reinforcing the lexicalized nature of *c'est* influencing both *ne*-retention and SD. In Near-NSs, the SD distribution is more equal

[116] NNS = near-native interlocutors in Lille.

between copula (10/14 SD) and non-copula contexts (10/17 SD), and similar results obtain for NSs (copula: 36/43 SD; non-copula 30/47 SD).

Regarding SD use in *ne*-retention contexts, the range across speaker groups is much higher (0% to 43%), though low token counts may partially explain the high percentage for near-native interlocutors in Lille. Since there are few examples of SD in *ne*-retention contexts, all 12 occurrences are listed here.

(1) a. *…les Le Pen ils ne sont pas aimés.* (ThF)
'the Le Pens (they) are not liked.'

b. *le niveau des professeurs d'anglais qui sont français il n'est pas bon.* (FrE)
'the level of English teachers who are French (it) isn't good.'

c. *les questions de didactique ou de pédagogie et cetera, ça n'existe pas*. (5L)
'the questions of didactics and pedagogy et cetera (that) does not exist.'

d. *parce que eux aussi ils n'utilisent pas les…certain temps.* (9L)
'because they too (they) do not use certain tenses.'

e. *l'équivalent ce n'est, ce n'est pas toujours évident.* (10L)
'the equivalent (it) is not, it is not always obvious.'

f. *mais les jeunes ici ils n'ont jamais les…* (1P)
'but young people here (they) never have the…'

g. *le français bientôt il n'existerait plus* (2P)
'French soon (it) would no longer exist.'

h. *Hamlet, qui était un acteur beaucoup plus expérimenté…il ne pouvait pas m'aider* (6P)
'Hamlet, who was a much more experienced actor…(he) could not help me.'

i. *et le lendemain, Farage il n'a pas pu répondre* (9P)
'and the next day, Farage (he) wasn't able to respond.'

j. *et ma mère elle n'avait pas de passeport* (SaE)
'and my mother (she) didn't have a passport.'

k. *notre point de vue ce n'est pas le seul* (SaE)
'our point of view (it) is not the only one.'

l. *et mon petit frère il n'habite pas loin aussi* (JeE)
'and my little brother (he) doesn't live far away also.'

It is possible that a categorical difference exists here between the native speakers in Lille, who produced no tokens of this kind of sentence (as Meisner found), and two of the bilingual speakers in Pau (and several of the Near-NSs), whose behavior indicates that these two elements can indeed co-occur, though in rare circumstances (as Villeneuve and Auger found).

Now let us compare the distribution of variable SD in all negative clauses, as Villeneuve and Auger (2013) did. In order to increase the accuracy of comparison with the latter study, tokens from the current study containing lexicalized *c'est* were excluded from Table 6-5.[117]

Table 6-5. Distribution of SD/negation contexts in Villeneuve & Auger (2013) and current study

	Villeneuve & Auger speakers		**Current study speakers**		
SD & negation patterns	Monolinguals	Bilinguals	NSs[118]	Near-NSs[119]	SA learners
1. no SD; *ne*-retention	11 (55%)	20 (63%)	18 (21%)	54 (43%)	18 (62%)
2. SD; *ne*-deletion	4 (20%)	7 (22%)	41 (48%)	39 (31%)	8 (28%)
3. no SD; *ne*-deletion	4 (20%)	4 (13%)	25 (29%)	22 (18%)	3 (10%)
4. SD; *ne*-retention	1 (5%)	1 (3%)	2 (2%)	10 (8%)	0 (0%)
Total tokens	20	32	86	125	29

Comparing the results across both studies, two important observations can be made. First, the Near-NSs and the SA learners match quite closely the distribution observed in Villeneuve and Auger's native speakers; not only does the same order of frequency obtain in all of these groups, but the percentages for each pattern also match to a large extent. However, the one group that does not follow the distribution obtained in Villeneuve and Auger's study is the NS group in the current

[117] The other excluded collocations that produce nearly categorical *ne*-deletion (*il y a*, *il faut*) cannot occur in subject doubling contexts.

[118] Includes the native interlocutor for the SA learners (two tokens).

[119] Includes the near-native interlocutors in Lille (13 tokens) and the near-native interlocutor for the SA learners (four tokens).

study. The first clear difference from this apparent outlier concerns the reversal of the percentages for the first two patterns. NSs in the current study produced considerably more instances of SD co-occurring with *ne*-deletion (e.g., *Ma mère elle parle pas français*) with comparatively fewer contexts of non-doubled subjects co-occurring with *ne*-retention (e.g., *Ma mère ne parle pas français*). This is the case even after 26 tokens of *c'est* co-occurring with SD and *ne*-deletion were excluded; furthermore, as revealed in Table 6-4, overall SD percentages in *ne*-deletion and *ne*-retention contexts for NSs do not differ greatly from the SD percentages found in the other groups. What this largely reveals for these NSs is simply a predominance of informal forms, clearly indicating their adoption of an informal style to a greater extent than learners in the current study as well as NSs in previous studies such as Villeneuve and Auger (2013); recall also Coveney's (2002) study indicating that some NSs were clearly quite conservative with *ne*-retention even when the conversation was expected to be informal.

The second observation concerns pattern 3 in Table 6-5 (non-doubled subjects co-occurring with *ne*-deletion, e.g., *Ma mère parle pas français*). NSs have a higher frequency of this pattern compared to all other groups. When there is no SD, NSs have much higher *ne*-deletion than all other groups. This helps explain the discrepancy found in the first two patterns, but only partly. The other variable that could explain this outlier may, in fact, be due to individual differences. One of the NSs, Fr, is responsible for 14 of the 25 tokens in pattern 3, but only one token in pattern 1. This bilingual speaker simply seems to be more comfortable than other NSs in dropping *ne* without doubling the subject. Further research would be needed, however, to determine whether Fr's distribution lies within the range of distributions of a larger sample of NSs (especially monolinguals), or if his speech patterns regarding these variables are an outlier to the extent that they may not be considered nativelike, at least from these speech samples.

6.3 Correlations of *ne*-deletion and SD: Group analyses

If we plot each speaker group's raw percentages of the use of the informal variants (*ne*-deletion and presence of SD)[120] we find an impressive symmetry: higher *ne*-deletion rates match almost perfectly with higher SD rates. In Figure 6-1, the speaker groups are organized left-to-right according to increasing SD percentage; in all but one case, the percentage of *ne*-deletion increases left-to-right as well, and the one difference (92.9% *ne*-deletion in SA interlocutors versus 93.7% in Pau NSs) is extremely marginal. (Note that 'NNS Lille' refers to the near-native interlocutors in Lille).

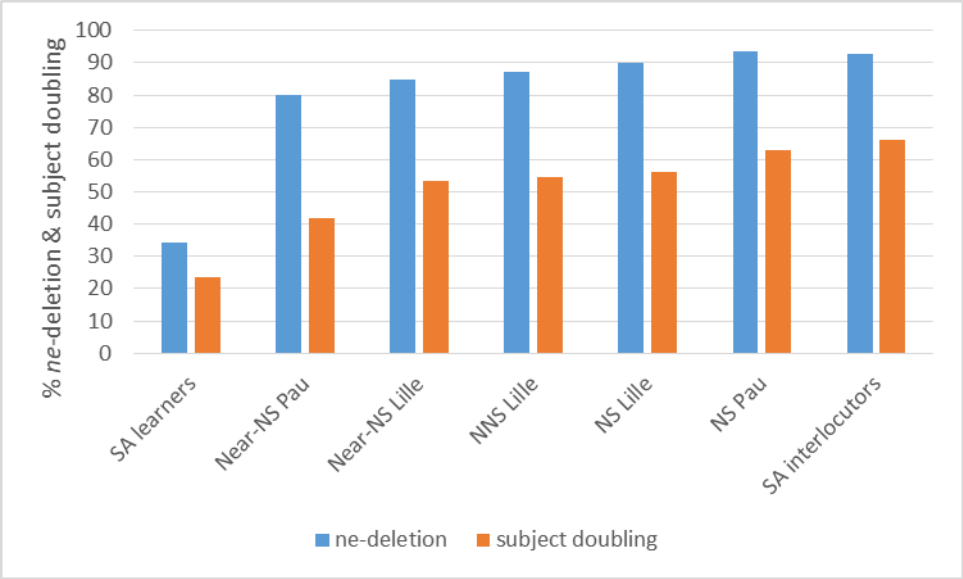

Figure 6-1. Ne-deletion plotted by increasing SD percentages across each speaker group

Ceiling effects and individual variation aside, the above figure demonstrates that *ne*-deletion and SD use are highly correlated across different proficiency levels. Expressing this relationship differently, Figure 6-2 presents SD rates as a percentage of *ne*-deletion rates.

[120] I emphasize that in this section, negation is expressed in terms of *ne*-deletion, in order to more easily visualize the concordance of informal variants.

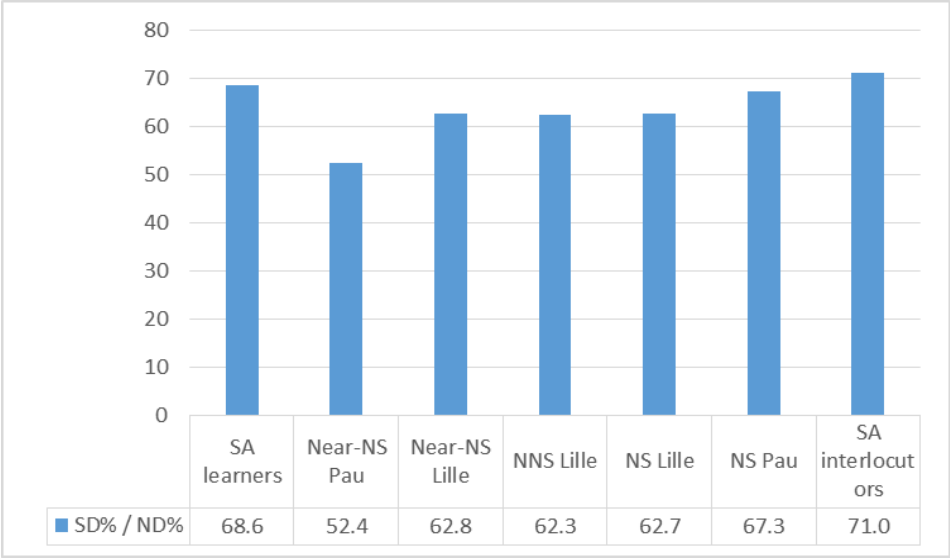

Figure 6-2. SD rates as a percentage of ne-deletion rates across all groups.

Across each group, the proportion of SD to *ne*-deletion is quite consistent. The largest outlier, at nearly two standard deviations lower than the mean (*SD* = 6.08; mean = 63.9%), Near-NSs in Pau, would be expected to have either higher SD, given their *ne*-deletion rates, or lower *ne*-deletion, given their SD rates. At the other end, the SA interlocutors are slightly greater than one standard deviation above the mean, suggesting that they have slightly higher *ne*-deletion in proportion to SD (or slightly lower SD in proportion to *ne*-deletion) than the overall average of all speaker groups. This concords with the "lag" in SD observed in some individual SA learners who otherwise have comparatively high *ne*-deletion.

Note that the correlation between these two variables is only very minimally supported by the direct interaction of these variables, as discussed in section 6.2. The number of tokens involving contexts for the presence/absence of both variables accounts for 11% of overall SD contexts (289/2674) and 6% of overall negation contexts (289/5109). Such tokens could potentially skew the overall proportions of SD and *ne*-retention. However, total SD usage only decreases by 0.2% when negation contexts are eliminated, and none of the speaker groups' percentages changes by

more than 0.5% in either direction. This is due, of course, to the fact that both types of negation contexts (*ne*-deletion and *ne*-retention) are removed, which eliminates SD and non-SD tokens respectively favored in each context. Total *ne*-retention, on the other hand, is slightly decreased when eliminating SD contexts (16.7% versus 17.9% overall), and each speaker group except for SA learners slightly decreases *ne*-retention here compared to overall rates (with the smallest change at 0.2% increase for SA learners and the largest change at 2.5% decrease for Near-NSs in Pau). This decrease is due to the fact that, as found in the *ne*-retention variationist analysis, lexical subject NPs strongly favor *ne*-retention; even though they appear in a small percentage of negation contexts, they make up the vast majority of SD contexts, decreasing the overall *ne*-retention rate when eliminated.

Thus, when the direct interaction of these two variables is removed, the overall proportions in the remaining data are essentially the same. Crucially, the L2 speakers, as a whole, pattern very similarly to native speakers; that is, there is very little asymmetry in the proportion of use of these variables between each group of L1 speakers and L2 speakers (across multiple proficiency levels), even though the raw percentages of each variable differ across groups and certain individual speakers show asymmetries. The next subsections will examine this individual variation in more detail for each of the three broad speaker groups (SA learners, Near-NSs, and interlocutors).

6.3.1 Correlations of *ne*-deletion and SD: SA learners

If we plot each learner's raw percentages of *ne*-deletion and subject doubling, we find that, for most SA learners, the variables are highly correlated: higher *ne*-deletion co-occurs with higher subject doubling, as Figure 6-3 illustrates. (Note again that *ne*-retention percentages have been converted to *ne*-deletion here, so that higher bars reflect more informal usage of both variables.)

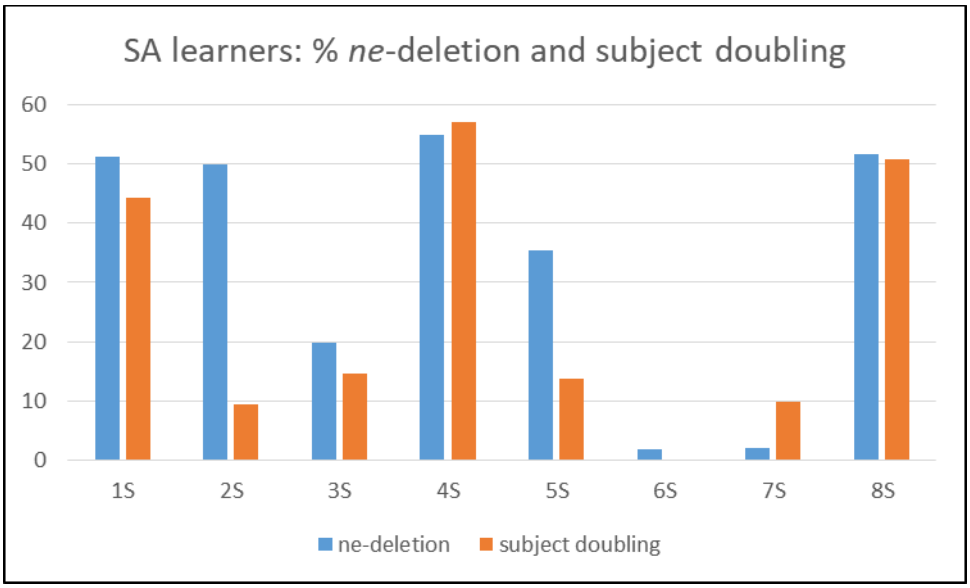

Figure 6-3. SA learners' use of ne-deletion and subject doubling

It is clear that higher *ne*-deletion and SD can be expected in informal styles compared to formal styles, even for learners at intermediate/pre-advanced levels. It should not be expected, however, that these two variables match each other so closely in terms of their occurrence as a percentage of total possible contexts, given that the two variables are conditioned by many different linguistic constraints, and again, the fact that one variable is defined in terms of its absence in informal styles (*ne*) while the other is defined in terms of its presence (SD).

At this point in their sociolinguistic development, the SA learners sampled in the current study can be separated into three groups, using Figure 6-3 as a reference.

1) Learners with low (or no) use of the informal variants (speakers 6S and 7S). At this stage, learners show little asymmetry with regard to these variables, suggesting that use of both informal variants may become incorporated into their speech at similar points later in the acquisition process. Perhaps unsurprisingly, these learners (along with 3S) had the lowest language security indices (as measured by their self-reported proficiency scores), which matches their sociolinguistic performance more closely than their *c*-test scores.

2) Learners with large asymmetries in use of the informal variants (speakers 2S and 5S). These two learners actually have the lowest (30 for 2S) and highest (46 for 5S) *c*-test scores of the entire group, whereas their language security indices are tied for second-highest in the group. At this stage in their sociolinguistic development, both of these learners clearly demonstrate knowledge and use of both informal variants; however, in both cases, *ne*-deletion is much more widespread than SD. Whether due to input clues or facilitation of one informal variant over another in their spoken production, these results suggest that more targetlike *ne*-deletion is acquired before more targetlike SD in informal contexts.
3) Learners with symmetrical high frequencies of *ne*-deletion and SD (speakers 1S, 4S, and 8S). These speakers' *c*-test scores are in the middle (38, 37, and 37, respectively) and their language security indices overlap with 2S and 5S, though 1S has the highest security index. These three learners appear to be as advanced as some highly proficient learners with regard to the use of both informal variants.

Whether these results suggest a developmental profile typical of most learners remains an open question, absent longitudinal data and a larger sample size. Nevertheless, it appears that use of the informal variant for both *ne*-deletion and subject doubling begins at similar points in time, and *ne*-deletion may be much more frequent than SD for some learners, but not the opposite—that is, learners with high use of SD also have high use of *ne*-deletion, but learners with high *ne*-deletion do not necessarily have high SD rates. Furthermore, it would be instructive to follow the trajectory of speaker 7S, the only learner with SD rates more than 2% higher than *ne*-deletion rates, in order to see whether (or at which point) *ne*-deletion "catches up," as the remaining developmental profiles predict.

6.3.2 Correlations of *ne*-deletion and SD: Near-NSs

Concerning the two Near-NS groups, some similarities with SA learners can be observed as well. Figures 6-4 and 6-5 chart the two variables for Near-NSs in Lille and Pau, respectively.

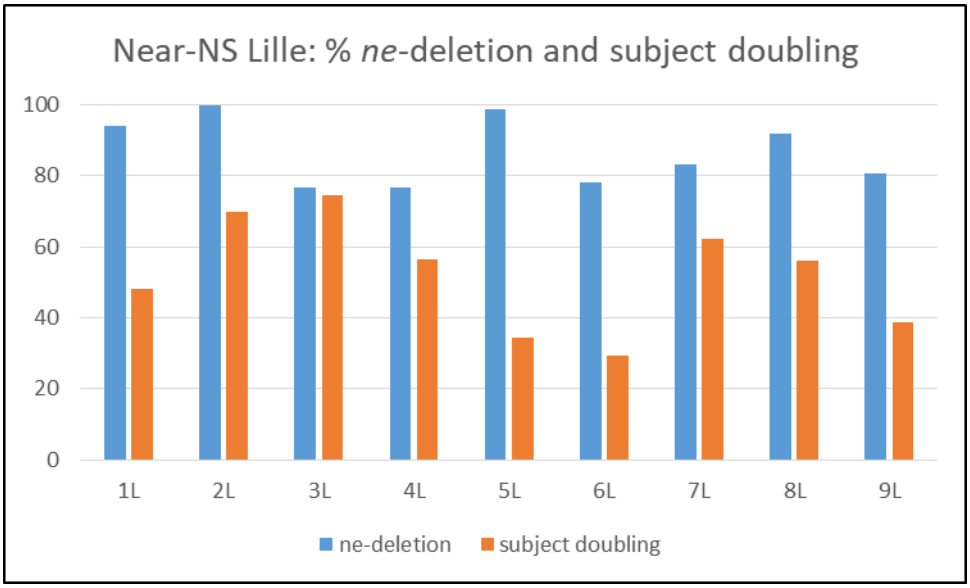

Figure 6-4. Near-NSs' use of ne-deletion and subject doubling (Lille)

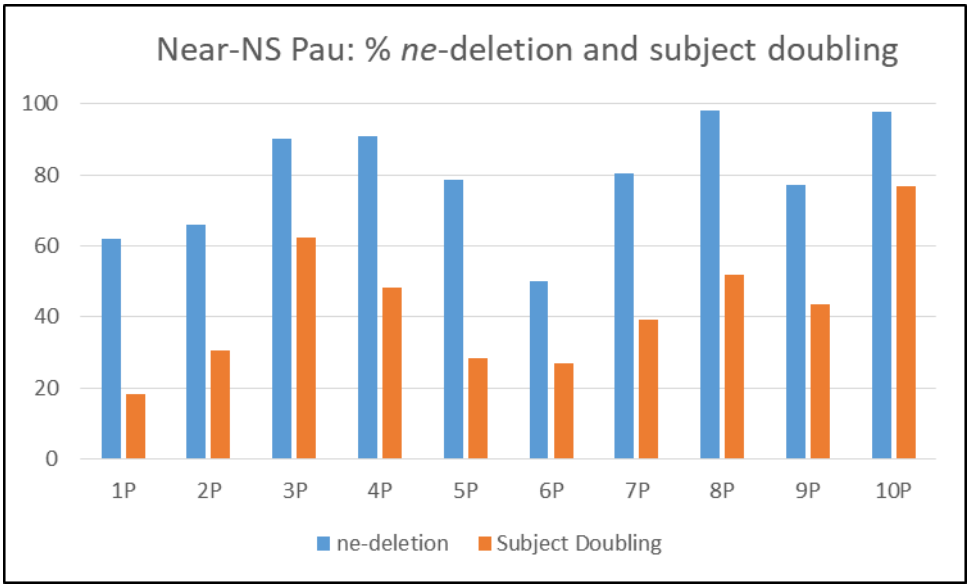

Figure 6-5. Near-NSs' use of ne-deletion and subject doubling (Pau)

An initial broad observation is that, regardless of site, every Near-NS has higher raw percentages of *ne*-deletion than SD. For Near-NSs in Lille, the inter-speaker *ne*-deletion ranges are lower (23.3%; low of 76.7% (4L) versus high of 100.0% (2L)) than the SD ranges (44.9%; low of 29.5% (6L) versus high of 74.4% (3L)). In Pau, the same pattern is observed, though higher inter-speaker ranges for both variables obtain: 49.0% range in *ne*-deletion (low of 49.3% (6P) versus high of 98.3% (8P)) compared with 58.5% range in SD (low of 18.2% (1P) versus high of 76.7% (10P)). There are also individual differences meriting mention. In Lille, speaker 3L has the highest SD and is essentially tied for the lowest *ne*-deletion (77.0% versus speaker 4L, 75.7%) in the group. Speaker 5L has nearly categorical *ne*-deletion but the second-lowest SD frequency. In Pau, there seem to be fewer extremes, though one would expect speaker 6P to have somewhat higher *ne*-deletion based on her SD usage.

In fact, the Near-NSs in Pau broadly illustrate the correlation between the two variables. When aligning individual speakers from lowest to highest SD percentage, corresponding increases in *ne*-deletion are obtained, with a general trendline highlighting this increase in Figure 6-6.

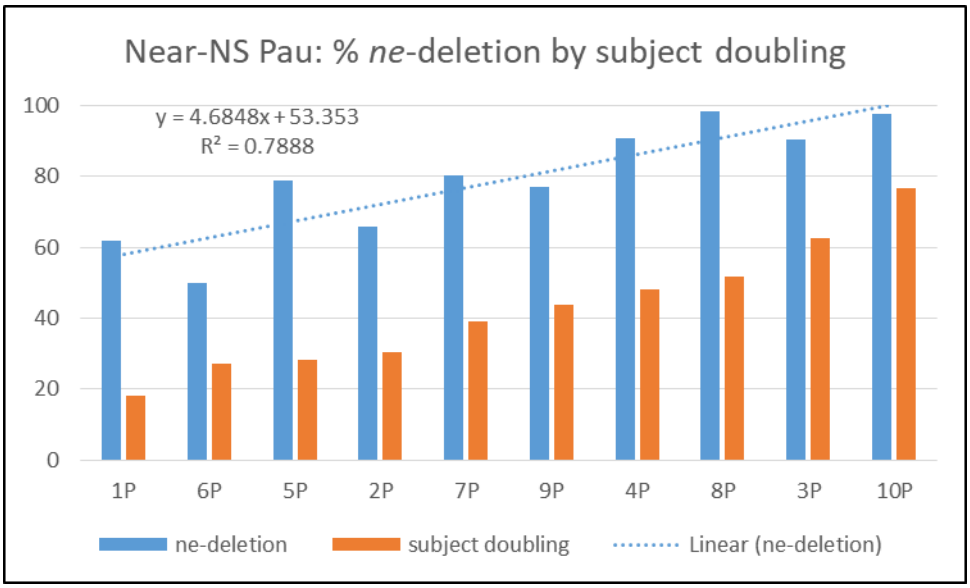

Figure 6-6. Ne-deletion plotted by increasing subject doubling: Near-NSs (Pau)

Due to the asymmetries observed in several of the Near-NSs in Lille, however, the trendline for this group is essentially flat, as Figure 6-7 indicates.

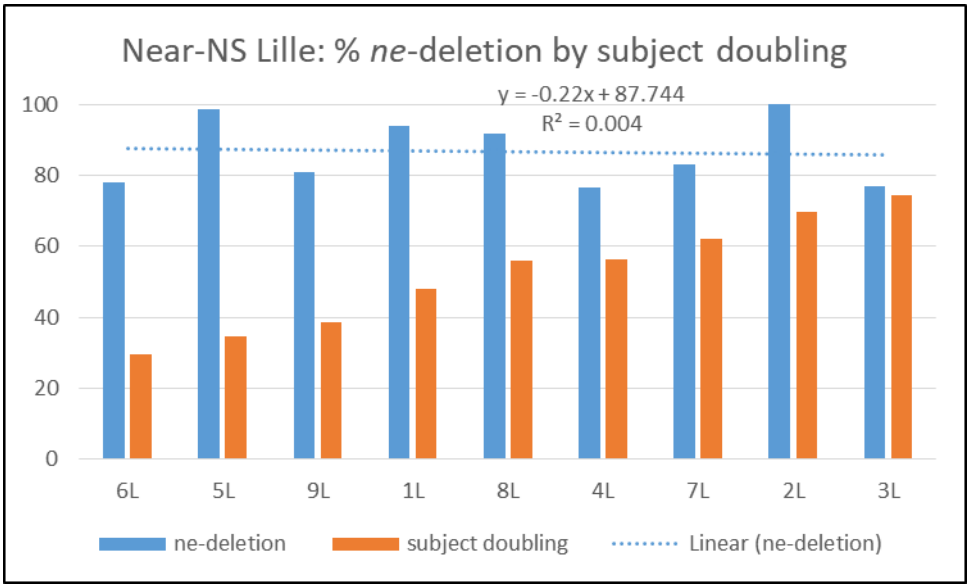

Figure 6-7. Ne-deletion plotted by increasing subject doubling: Near-NSs (Lille)

Again, if *ne*-deletion and SD are more closely correlated, we would expect comparatively higher *ne*-deletion for speaker 3L and comparatively lower *ne*-deletion for several speakers (6L, 5L, 9L, 1L, and 8L). For these speakers, distribution of these variables may not reflect nativelike patterns.

6.3.3 Correlations of *ne*-deletion and SD: Interlocutors

Concerning individual variation in the NS and near-native interlocutors for each learner group, breaking down the use of these variables does reveal the same broad patterns as for the Near-NS groups: raw *ne*-deletion percentages are higher than SD percentages for each speaker (Figure 6-8).

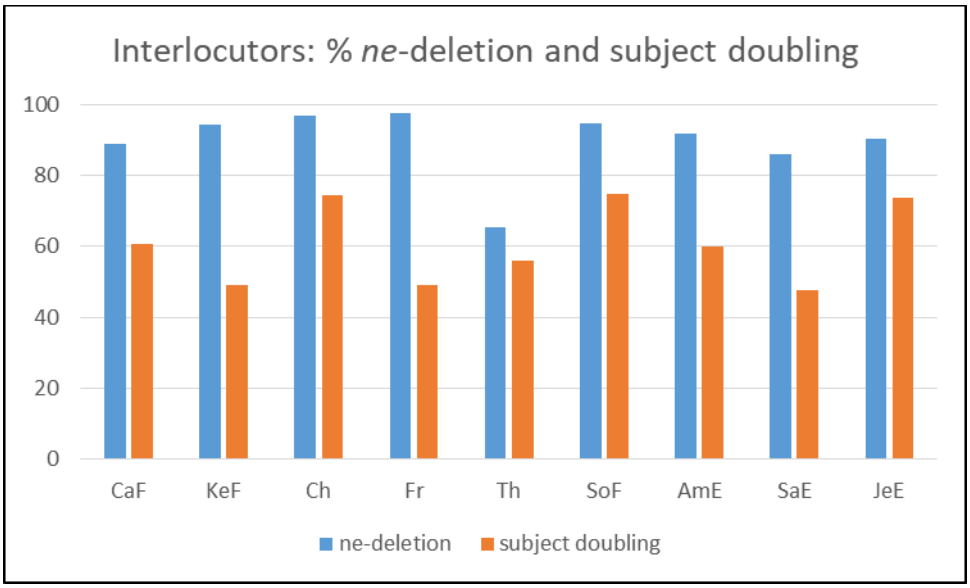

Figure 6-8. Interlocutors' use of ne-deletion and subject doubling (all sites)

Two observations can be made from these results. First, Th is a clear outlier, given his comparatively low *ne*-deletion rates, though these rates may be explained by one outlier conversation in an otherwise small sample (see section 4.4.3.1 for details). One would expect that, if he were adopting a more conservative speaking style and more carefully monitoring his speech, his SD rates would not be as high. This is likely indicative of the fact that *ne* is a more highly marked item in stylistic variation, whereas SD (especially with the *c'est* structure) is comparatively less marked and less likely to be consciously monitored. Second, in this group, even with NSs and near-natives combined, higher SD only slightly correlates with higher *ne*-deletion, as Figure 6-9 illustrates.[121]

[121] Removing speaker Th from this group allows for a slightly better fit of the trendline ($R^2 = 0.1016$) but produces even less of an increase in the *ne*-deletion slope ($y = 0.4389x + 90.313$).

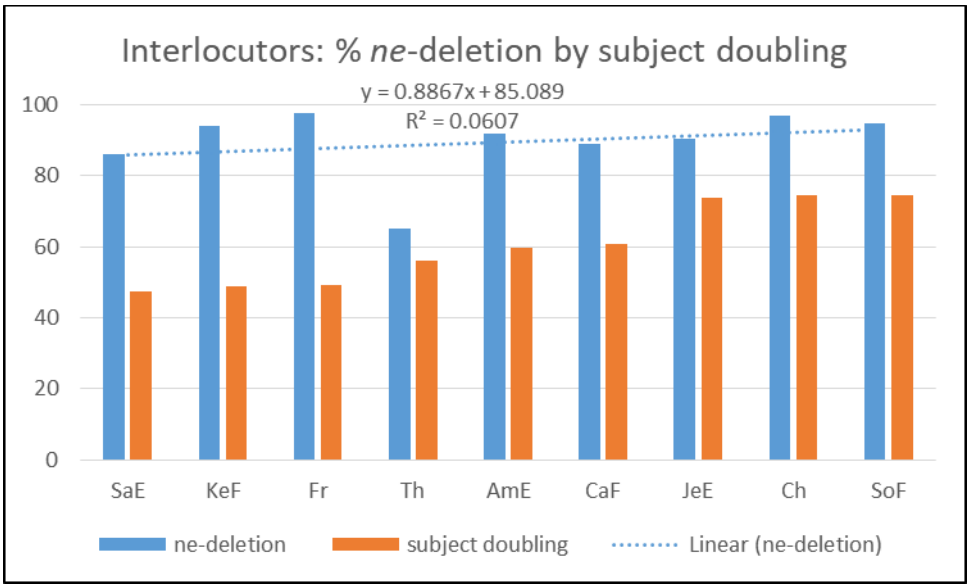

Figure 6-9. Ne-deletion plotted by increasing subject doubling: Interlocutors (all sites)

Aside from speaker Th, these interlocutors generally appear to be approaching ceiling with their *ne*-deletion rates, with only small individual variation (low = 86.0%; high = 97.5%). Thus, a higher SD rate appears to have a comparatively small effect on most speakers' *ne*-deletion rates.

6.3.4 *Ne*-deletion and SD: Comparisons with Coveney (2005)

In his article on subject doubling, Coveney (2005: 106) makes a brief comparison of SD and *ne*-deletion rates in his L1 French speakers, and a comparison with my corpus merits attention here. As in my data, Coveney finds a "moderately good" correlation between the two variables, with a scatterplot for all 30 speakers. Figure 6-10 reproduces Coveney's chart converted in terms of *ne*-deletion and SD, as I have done for my data in this section. That is, the most formal styles will appear in the lower-left corner and the most informal styles will appear in the upper-right.

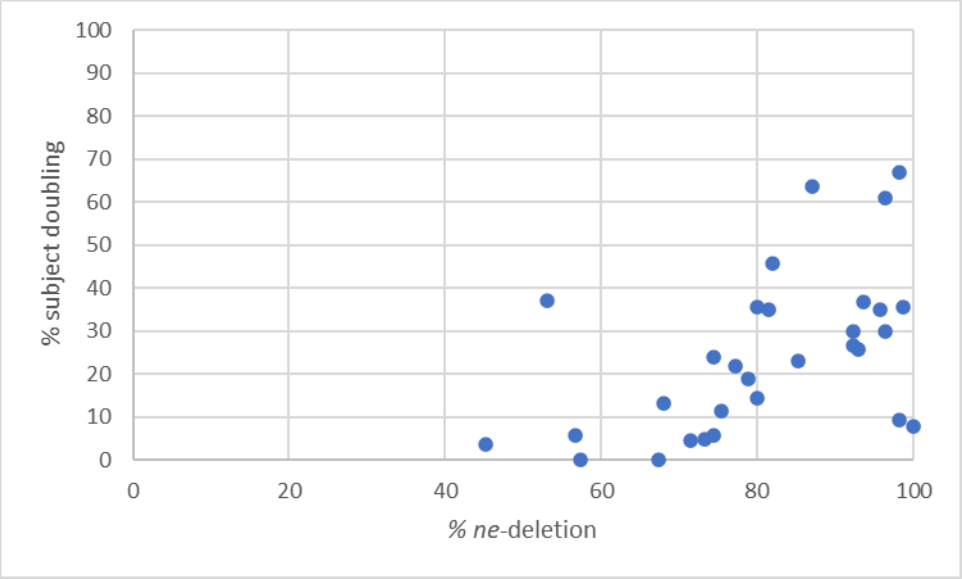

Figure 6-10. Ne-deletion and SD plots in Coveney (2005)

Coveney also provides the ages for each speaker in his chart, noting that most of his younger informants are in the lower-left (for my chart) and most mid-age speakers are in the upper-right.

For my corpus, Figure 6-11 provides the scatterplot of the 36 speakers, with colors coded for each speaker group.

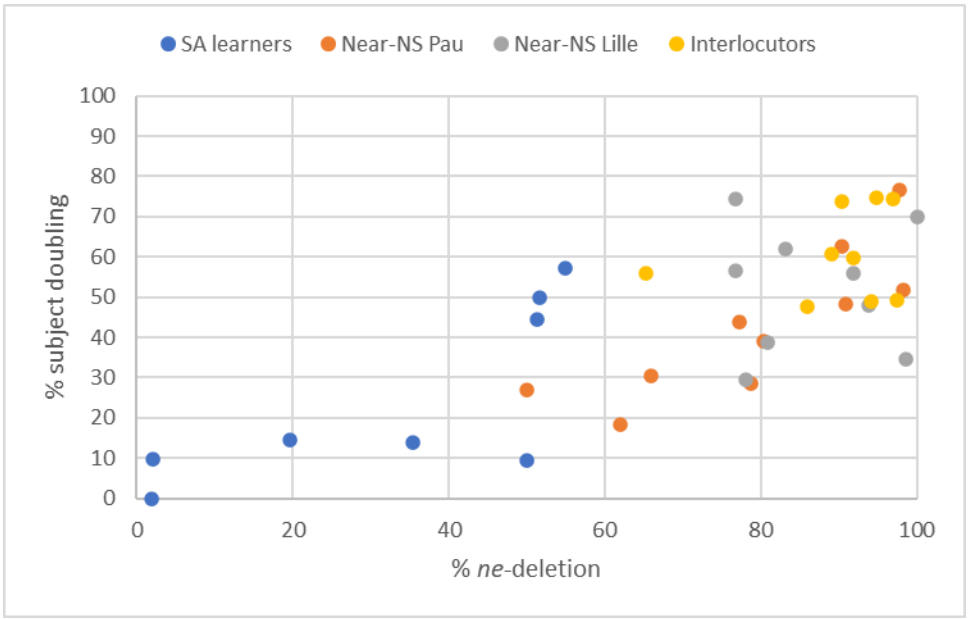

Figure 6-11. Ne-deletion and SD plots in the current study

As expected, SA learners dominate the lower-left of the chart, while most Near-NSs are found with the native and near-native interlocutors. Coveney's native speakers tend to be more formal with their use of SD in particular, with a third of his speakers lower than the lowest Near-NS SD rate (18%) in my corpus. Coveney also has two speakers near the lower-right corner, with near-categorical *ne*-deletion but low SD rates (8-9%), whereas my native and near-native speakers show a stronger correlation between the two variables. Finally, Coveney notes that most of his speakers who have above-average *ne*-deletion (higher than 81%) also have higher than average SD (above 25%). As discussed earlier in section 6.3, in my corpus this pattern applies to most of the SA learners (with two outliers) and all but three of the Near-NSs.

6.3.5 Correlations of *ne*-deletion and SD (and beyond): Future considerations

To close this section, we can identify many questions for future consideration based on the observations made for these correlations. If we assume that the NSs in the current study (which includes, of course, bilingual speakers considered as L1 French speakers) are representative of (at least) Hexagonal French native speakers, then we have a baseline for a proportion of these two variables in an informal style. When analyzing the acquisition of sociolinguistic variation, we can certainly calculate to what extent learners deviate from the range of proportions of the native speakers. However, just how perceptible are such deviations when examined in the context of an oral production task? Are any asymmetries, with regard to overall proportions of these variables, perceptible to an interlocutor and/or another observer? For example, would bilingual Th's comparatively high SD rates, given his high *ne*-retention, be noticed (that is, seem unusual) by any native speakers? At what threshold would particular asymmetries identify a speaker as "outside the norm" or "non-native" in some way? Or would highly proficient French users (native as well as non-native) react more strongly to more specific examples of both variables in interaction (e.g.,

presence or absence of *ne* and presence or absence of SD in a context where both are possible, or other specific utterances that NSs/Near-NSs would not produce in informal contexts (such as repeated use of *ce n'est pas* versus *c'est pas*))? How strongly do NSs react to the tendency for non-natives (at different proficiency levels) to overuse neuter *ce/ça* when doubling [-animate] NPs? Such perception tasks, based on the corpus obtained in the current study, will form the basis of future studies.

6.4 Conclusions from the current study: Pedagogical implications

I now turn to implications of the conclusions drawn from the current study, beginning with pedagogical implications in this section, followed by methodological implications in section 6.5. As the discussions of pedagogical approaches to the use of *ne*-retention (section 4.3) and SD (section 5.3.2) have shown, teachers and textbooks alike do not necessarily model the use of these two variables in the same way that they are used in authentic informal styles.

Contemporary college classroom curricula increasingly emphasize the speaking aspect of language learning as opposed to more traditional curricula emphasizing comprehension and production of written forms of language. Since the speaking aspect encompasses multiple styles involving manipulation of numerous sociolinguistic variables, Sax (2003) advocates a pedagogical approach based on the fact that learners begin incorporating aspects of informal spoken French on their own as they interact with target language communities, and that this learner speech may be non-targetlike due to lack of stylistic awareness (e.g., overuse of informal forms in formal contexts and vice-versa). Therefore, Sax argues that metalinguistic awareness should be part of the (instructed) language acquisition process. Since learners are likely to (eventually) use informal forms of the L2 in interactions with the target language community outside the classroom, teachers might as well instruct them on the social implications of their choices, rather than "ignoring" the

issue. By way of example, Sax suggests that learners be exposed to informal forms in the first year of instruction; by the second year, learners should be able to identify multiple aspects of stylistic variation and determine the likely formality of a speech sample; in the third and fourth years, learners practice adopting different styles through the use of role plays. From a case study examining *ne*-retention in a single learner, van Compernolle (2019) also advocates for explicit metalinguistic instruction alongside the use of contextual exemplars for creating a repertoire of specific lexicogrammatical templates (rather than an abstract "drop *ne* in informal speech" suggestion) in modeling sociolinguistic variation for learners. The current study shows that at least some of the least proficient learners might benefit from some type of metalinguistic instruction *before* their arrival in the target language community, rather than, as some SA learners confided concerning *ne*-retention, learning about sociolinguistic variation well after beginning their study-abroad program.

As Sax does, Fonseca-Greber (2000) likewise stresses that teachers would do well to clearly distinguish between written and spoken varieties, making it clear whether each structure presented is used primarily in a spoken or written (or formal spoken) style. Regarding *ne*-retention, for example, Fonseca-Greber suggests teaching *ne* as it is taught to L1 French speakers, as something that must be added to the post-verbal negator in writing or in formal speech, which is the opposite of how negation is usually introduced (with *ne* as a necessary element of bipartite negation which may, according to the teacher or textbook, sometimes be dropped in more casual speech). A more "controversial" suggestion, in her words, is Fonseca-Greber's approach to teaching inflectional prefixes (subject clitics), where the traditional *je/tu/il/elle/on/nous/vous/ils/elles* paradigm is ignored and the strong pronouns *moi/toi/*etc. are taught as the true, personal pronouns (in the same way that L1 French children acquire the paradigm), and where learners

acquire the pragmatic competence for understanding when to include or omit the subject clitics. The current study shows that even the least proficient learners have incorporated at least minimal use of the *moi, je…* sequence as a turn-taking strategy or to signal some sort of contrast; however, doubling with other strong pronouns is nearly non-existent at this stage. It would be instructive to see whether Fonseca-Greber's approach would "speed up" learner development of doubling with the strong pronoun paradigm.

As for doubling with lexical NPs, learners who have minimally achieved conversational competence and who wish to integrate into the target-language community may benefit from the pedagogical norm advocated by Ossipov (2002) for SD, in which neuter *ce* can simplify the gender/number computations necessary in marking verbal agreement. The results of the current study show that intermediate-advanced learners (SA learners) make almost exclusive use of this clitic when producing doubled subjects with [-animate] nouns, and Near-NSs highly favor it over personal subject pronouns in this context as well. Thus a good starting strategy for these learners, especially in contexts when topic, focus, or contrast needs to be established, would be to double with personal pronouns the subjects that have natural gender and double with *ce/ça* all other subjects.

As for near-native speakers whose goal is to attain nativelike proficiency, regardless whether their actual baseline SD percentages overlap with native speaker production, such learners would do well to focus on the use of personal pronouns for doubling in ambiguous contexts. This context is where the smallest overlap obtains across near-native and native speakers. Despite presenting a challenge in the online computation of gender, learners who pattern like NSs in the subject doubling distribution of this context would eliminate a particular asymmetry that persists in the overall grouping of native versus non-native speakers. Whether this asymmetry in non-

natives could be detected by native speakers in an authentic interaction or in a controlled experiment is an open question as addressed in section 6.3.3 (would NSs ever rely on this asymmetry as a cue to the non-nativeness of their interlocutor, all other linguistic performance being equal?). The current study suggests that nativelike performance by near-natives in this context is possible in principle (echoing such findings as those by Donaldson (2011a, 2011b) for left- and right-dislocation by near-natives), though more data would make this conclusion (and possible re-analysis of what constitutes a highly proficient near-native speaker) more robust.

Less clearly established is the symmetry in distribution of *ne*-deletion with non-doubled subjects. Though the Near-NSs (and SA learners, for that matter) pattern similarly to the NSs sampled in at least one previous study (Villeneuve & Auger, 2013), the NSs in the current study do not match the same distribution with respect to this pattern of interaction. This result may be explained either by the fact that individual differences produced one outlier (where the speaker's bilingual status may play a role), or that the possible distribution of this particular pattern in NSs is more varied than what has been previously obtained in a similarly small sample of NSs. Learners, then, would need to cautiously interpret any observations they may obtain from this distribution in such an interlocutor, whereas they can reasonably conclude, based on input from larger samples of NS speech, that informal French generally favors the co-occurrence of SD and *ne*-deletion while strongly disfavoring the co-occurrence of SD and *ne*-retention. Again, higher frequencies of the latter pattern (8%) are obtained in my Near-NSs compared with NS frequencies (between 2-5%); whether this higher frequency creates an asymmetry noticed by NSs, or whether any resulting asymmetry in learner speech is detectable in authentic interactions compared to a controlled experiment, also remains an open question.

6.5 Conclusions from the current study: Methodological implications

As outlined in section 2.9.1, oral production tasks involving measures of sociolinguistic performance used in many studies are inconsistent with respect to the language background status of the interviewer/interlocutor, and the choice of interlocutor is, understandably, often made out of convenience. However, when eliciting oral production from learners in order to make observations about learner sociolinguistic performance, it may be preferable to consider authentic native speakers as interlocutors, especially with lower-proficiency learners, as such interlocutors may allow researchers to see what learners may actually be capable of producing when conditions are more ideal for eliciting more nativelike speech. High-proficiency learners appear to be less influenced by the interlocutor L1 factor, but the fact that this group still shows some sensitivity with regard to SD use indicates that the selection of interlocutors must still be considered with care when designing oral production tasks. For future studies, it remains to be seen whether other sociolinguistic variables show sensitivity in the same ways as *ne*-retention and SD, whether different linguistic domains (e.g., phonology, syntax, lexicon) are influenced by the same factor, and whether these domains would reveal similar trends in sensitivity for some learner groups but not others.

6.5.1 Limitations in the current study

In light of the considerations for other researchers outlined above, I briefly acknowledge here some limitations inherent in the current study. The most obvious limitations involve the number and selection of participants. Regarding number, it would have been ideal to recruit at least 10 L2 speakers for each group. Though the goal was 10 speakers, the availability of all contacted participants became less manageable due to the necessity of coordinating schedules for two to three conversation partners (in addition to myself). The challenges in coordinating

schedules also impacted the selection of participants, especially interlocutors for the Near-NS groups. It was necessary to recruit a third bilingual in Pau and a second interlocutor for both the NS and near-native conversation partner; reducing the inherent variability in having more interlocutors would make the observations from the Near-NSs groups' behavior more robust. As for selection of Near-NSs, the ranges in age and length of residence, as well as conversational dyads mismatched for age and gender, reduce the generalizability of the results. In addition, one could also question the inclusion of at least one participant as a near-native speaker (6L), due to low language security and a high frequency of grammatical errors relative to other Near-NSs.

Regarding other tasks performed by learners in the current study, it would have been useful to consider other measures for evaluating proficiency beyond the *c*-test and AJT, since there were not strong correlations between these measures and learners' informally observed proficiency in the production tasks. It would also have been helpful to revise the language background questionnaire to include more details on learners' interactions with native speakers in the community. A questionnaire more similar to that found in Nagy et al. (2003) would make comparisons with that learner group more robust while validating the observation that increased interaction with NSs correlates closely with more nativelike sociolinguistic behavior. Furthermore, in the follow-up debriefings that I conducted, it would have been instructive to include a qualitative analysis of subject doubling in addition to the *ne*-retention variable that I explicitly discussed with each learner.

As with all cross-sectional studies, the results from this study were obtained at a single point in time (albeit spread over multiple conversations). It would be instructional to observe the use of sociolinguistic variables, and to measure for an interlocutor effect, in the same SA learners at a later point in their L2 development, or to make such longitudinal observations in learners

undertaking a longer study-abroad program, as other research groups (e.g., Regan, Howard, & Lemée, 2009) have done. Furthermore, the design of the current study does not allow for observations of the same SA learners in repeated interactions with the same interlocutors; use of sociolinguistic variables may change after familiarity has been established with such interlocutors.

Finally, the conclusions drawn from my results could benefit from more robust quantitative analyses. These include specific factor groups coded in the variationist analyses (such as the relatively subjective determination of emphasized negation) as well as the determination of SD contexts due to the challenges of separating cases of dislocation from subject doubling.

Despite these limitations, I consider the current project to be an important advancement in the state of knowledge of the fields of sociolinguistics and second language acquisition. The results of this study shed more light on the behavior of certain less commonly studied learner groups, such as near-native speakers, while analyzing usage of less commonly studied variables in learners, such as subject doubling, all examined in the context of an uncommonly manipulated social dimension that can have implications on how sociolinguistic competence is measured in learners.

6.6 Revisiting a language learner's questions

I shall conclude by returning to questions posed in the introduction of this study, viz.: Why did I feel like my French "flowed" more naturally with native speakers but often required more of an effort with non-native speakers, even those who were already fluent in French? Was this the case for other learners of French?

The data obtained in the current study can address these questions. On at least two measures of sociolinguistic performance (i.e., two variables conditioned by stylistic differences), the interlocutor's language background appears to condition the degree to which this performance resembles native speaker patterns, an effect observed in pre-advanced learners and at least some

highly proficient learners. My level of proficiency at the time I began posing these questions was likely closer to that of the SA learners rather than that of the Near-NSs sampled in the current study. As with these learners, the interlocutor effect seems to have influenced how I approached the task of communicating in my second language. Though other learners may not be as conscious of changes in their language patterns when speaking to interlocutors of different language backgrounds, I felt that my ability to sound more nativelike was facilitated with some speakers and inhibited with others (such as with my interlocutor in my Oral Proficiency Interview), at least in initial interactions with such speakers. The current study provides evidence that these changes in language patterns exist for other learners as well, and the results indicate that such changes are detectable and quantifiable.

As many previous investigations on language use have demonstrated, however, more questions can be raised than can be answered in this study alone. In addition to the questions already posed in this final chapter concerning perceptions of asymmetry, one can question how extensively the interlocutor effect conditions the language acquisition process (and language, more generally). Is this effect detectable in other sociolinguistic variables? Are other linguistic domains more, or less, sensitive to this effect? (Are more salient forms, such as informal vocabulary, subject to the same effect?) In which domains might this effect be more detectable on highly proficient learners? How might this effect reveal itself in native speakers of languages other than English, or in learners of languages other than French? These questions are only the beginning of what constitutes a rich area for future research probing the limits on, and expectations for, how language is used in its social context.

Bibliography

Abu-Rabia, S., & Kehat, S. (2004). The critical period for second language pronunciation: is there such a thing? *Educational Psychology*, *24*(1), 77–97.

Adams, R., Nuevo, A. M., & Egi, T. (2011). Explicit and implicit feedback, modified output, and SLA: Does explicit and implicit feedback promote learning and learner-learner interactions? *The Modern Language Journal*, *95* (Supplement 1), 42–63.

Adamson, H. D. (1988). *Variation theory and second language acquisition*. Georgetown University Press.

Adamson, H. D. (2009). *Interlanguage variation in theoretical and pedagogical perspective*. Routledge.

Adamson, H. D., & Regan, V. (1991). The acquisition of community speech norms by Asian immigrants learning English as a second language. *Studies in Second Language Acquisition*, *13*, 1–22.

Allen, H.W., & Herron, C. (2003). A mixed methodology investigation of the linguistic and affective outcomes of summer study abroad. *Foreign Language Annals*, *46*, 377–401.

Antes, T. (1995). The move towards topic prominence in spoken French: Grammatical and pedagogical repercussions. *Michigan Academician*, *27*, 503–14.

Ariew, R. & Nerenz, A. (1989). *C'est à dire*. Heinle & Heinle.

Armstrong, N. (1996). Deletion of /l/ in French: Linguistic, social and stylistic factors. *Journal of French Language Studies*, *6*, 1–21.

Armstrong, N. (2002). Variable deletion of French *ne*: A cross-stylistic perspective. *Language Sciences*, *24*, 153–173.

Arnett, C. (2013). Syntactic gains in short-term study abroad. *Foreign Language Annals*, *46*(4), 705–712.

Arthur, B., Weiner, R., Culver, M., Lee, Y.J., & Thomas, D. (1980). The register of impersonal discourse to foreigners: verbal adjustments to foreign accent. In D. Larsen-Freeman (Ed.), *Discourse analysis in second language research* (pp. 111–124). Newbury House.

Arva, V., & Medgyes, P. (2000). Native and non-native teachers in the classroom. *System*, *28*, 355–372.

Ashby, W. (1976). The loss of the negative morpheme, *ne*, in Parisian French. *Lingua*, *39*, 119–37.

Ashby, W. (1980). Prefixed conjugation in Parisian French. In H.J. Izzo (Ed.), *Italic and Romance: Linguistic studies in honor of Ernst Pulgram* (pp. 195–207). John Benjamins.

Ashby, W. (1981). The loss of the negative particle *ne* in French: A syntactic change in progress. *Language*, *57*, 674–687.

Ashby, W. (1982). The drift of French syntax. *Lingua*, *57*, 29–46.

Ashby, W. (1984). The elision of /l/ in French clitic pronouns and article. In E. Pulgram (Ed.), *Romanitas: Studies in Romance linguistics* (Michigan Romance Studies 4) (pp. 1–16). University of Michigan Press.

Ashby, W. (1988). The syntax, pragmatics, and sociolinguistics of left- and right- dislocations in French. *Lingua*, *75*, 203–229.

Ashby, W. (2001). Un nouveau regard sur la chute du *ne* en français parlé tourangeau: S'agit-il d'un changement en cours? *Journal of French Language Studies*, *11*, 1–22.

Audibert-Gibier, M. (1992). L'accord du participe passé en français parlé. *Langage et société*, *61*, 7–30.

Auger, J. (1991). *Variation and syntactic theory: Agreement-marking vs. dislocation in Québec Colloquial French* [Conference paper]. NWAVE 20, Georgetown University, Washington, D.C.

Auger, J. (1994). *Pronominal clitics in Québec colloquial French: A morphological analysis* (Publication No. 9520998) [Doctoral dissertation, University of Pennsylvania]. ProQuest Dissertations Publishing.

Auger, J. (1998). Le redoublement des sujets en français informel québécois : une approche variationniste. *Canadian Journal of Linguistics*, *43*(1), 37–63.

Auger, J. (2003). Le redoublement des sujets en picard. *Journal of French Language Studies*, *13*, 381–404.

Auger, J., & Valdman, A. (1999). Letting students hear the diverse voices of francophony. *Modern Language Journal*, *83*(3), 403–412.

Auger, J., & Villeneuve, A.-J. (2008). *Ne* deletion in Picard and in regional French. In M. Meyerhoff & N. Nagy (Eds.), *Social lives in language – Sociolinguistics and multilingual speech communities* (pp. 223–247). John Benjamins.

Auger, J., & Villeneuve, A.-J. (2010). La double expression des sujets en français saguenéen : étudie variationniste. In W. Remysen & D. Vincent (Eds.), *Hétérogénéité et homogénéité dans les pratiques langagières* (pp. 67–85). Les Presses de l'Université du Québec.

Avanzi, M. (2012). *L'interface prosodie/syntaxe en français : Dislocations, incises et asyndètes* (Gramm-R: études de linguistique française, no. 13). Peter Lang.

Ayres-Bennett, W. (1994). Negative evidence: Or another look at the non-use of *ne* in 17th century French. *French Studies*, *48*(1), 63–85.

Ayres-Bennett, W. (2005). *A History of the French language through texts.* Routledge.

Ball, R. (2000). *Colloquial French grammar: A practical guide.* Blackwell.

Bardovi-Harlig, K. (1999). Exploring the interlanguage of interlanguage pragmatics: A research agenda for acquisitional pragmatics. *Language Learning*, *49*, 677–713.

Bardovi-Harlig, K. (2012). Pragmatics in SLA. In S. M. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 147–162). Routledge/Taylor Francis.

Bardovi-Harlig, K. (2013). Developing L2 pragmatics. *Language Learning*, *63* (Supplement 1), 68–86.

Barnes, B. K. (1985). *The pragmatics of left-detachment in spoken Standard French*. John Benjamins.

Bayley, R. (2005). Second language acquisition and sociolinguistic variation. *Intercultural Communication Studies*, *14*, 1–13.

Bayley, R., & Regan, V. (2004). Introduction: The acquisition of sociolinguistic competence. *Journal of Sociolinguistics*, *8*, 323–338.

Bayley, R., & Tarone, E. (2011). Variationist perspectives. In S. M. Gass & A. Mackey (Eds.), *Handbook of second language acquisition* (pp. 41–56). Routledge.

Beauchemin, N., & Martel, P. (1977). *Recherches sociolinguistiques dans la région de Sherbrooke : Échantillon de textes libres III. Document de travail N. 10*. Université de Sherbrooke, Sherbrooke, Québec, Canada.

Beaulieu, L., & Balcom, P. (1998). Le statut des pronoms personnels sujets en français acadien du nord-est du Nouveau-Brunswick, *Linguistica Atlantica*, *11*, 1–27.

Beebe, L. (1977). The influence of the listener on code-switching. *Language Learning*, *27*, 331–339.

Beebe, L. (1985). Choosing the right stuff. In S. M. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 104–144). Newbury House Publishers.

Beebe, L., & Zuengler, J. (1983). Accommodation theory: An explanation for style shifting in second language dialects. In N. Wolfson & E. Judd (Eds.), *Sociolinguistics and language acquisition* (pp. 195–213). Newbury House Publishers.

Beeching, K. (2001). Repair strategies and social interaction in spontaneous spoken French: The pragmatic particle *enfin*. *Journal of French Language Studies*, *11*, 23–40.

Beeching, K. (2004). Pragmatic particles—polite but powerless? Tone-group terminal *hein* and *quoi* in contemporary spoken French. *Multilingua*, *23*, 61–84.

Beeching, K. (2011). The translation equivalence of *bon*, *enfin*, *well* and *I mean. Revue Française de Linguistique Appliquée*, *XVI*(2), 91–105.

Béguelin, M.-J., & Corminboeuf, G. (2017). *Ou comme ça*, *machin* et autres marqueurs d'indétermination dans les listes. *Discours*, *20*. https://doi.org/10.4000/discours.9275

Bell, A. (1984). Language style as audience design. *Language in Society*, *13*(2), 145–204.

Belletti, A. (1999). Italian/Romance clitics: structure and derivation. In H. van Riemsdijk (Ed.), *Clitics in the languages of Europe* (pp. 543–579). Mouton de Gruyter.

Benke, E., & Medgyes, P. (2005). Differences in teaching behavior between native and non-native speaker teachers: As seen by the learners. In E. Llurda (Ed.), *Nonnative language teachers: Perceptions, challenges and contributions to the profession* (pp. 195–216). Springer.

Berrendonner, A., & Béguelin, M.-J. (2020). Le verbe *c'est*. *Langue française*, *205*, 51–64.

Bertrand, R. & Chanet, C. (2005). Fonctions pragmatiques et prosodie de 'enfin' en français spontané. *Révue de Sémantique et Pragmatique*, 41–68. https://hal.archives-ouvertes.fr/hal-00353742

Biber, D. (1995). *Dimensions of register variation: A cross-linguistic comparison*. Cambridge University Press.

Biers, K. (2014). *Gender and interlocutor effects in French* (Publication No. 3631268) [Doctoral dissertation, Indiana University Bloomington]. ProQuest Dissertations Publishing.

Birdsong, D. (1992). Ultimate attainment in second language acquisition. *Language*, *68*, 706–755.

Blanche-Benveniste, C. (1994). Quelques caractéristiques grammaticales des « sujets » employés dans le français parlé des conversations. In M. Yaguello (Ed.), *Subjecthood and subjectivity: the status of the subject in linguistic theory* (pp. 77–107). Ophrys.

Blanche-Benveniste, C. (1997). *Approches de la langue parlée en français.* Ophrys.

Blasco-Dulbecco, M. (1999). *Les dislocations en français contemporain.* Champion.

Blondeau, H. & Nagy, N. (1998). Double marquage du sujet dans le français parlé par les jeunes anglo-montréalais. In J. Jensen & G. Van Herk (Eds.), *Actes du congrès annuel de l'Association canadienne de linguistique* (pp. 59–70). Cahiers Linguistique d'Ottawa.

Bourhis, R., Giles, H., Leyens, J.-P., & Tajfel, H. (1979). Psycholinguistic distinctiveness: Language divergence in Belgium. In H. Giles & R.N. St. Clair (Eds.), *Language and social psychology* (pp. 158–185). Basil Blackwell.

Bowles, M., Adams, R., & Toth, P. (2014). A comparison of L2-L2 and L2-heritage learner interactions in Spanish language classrooms. *The Modern Language Journal*, *98*(2), 497–517.

Bresnan, J., & Mchombo, S. A. (1987). Topic, pronoun, and agreement in Chichewa. *Language*, *63*(4), 741–782.

Campese, S. (2015, April 14). "'La France, elle est' : Effet de style ou faute de français ?" *Projet Voltaire*. https://www.projet-voltaire.fr/culture-generale/la-france-elle-est-effet-style-faute-francais/

Campion, E. (1984). *Left dislocation in Montreal French* (Publication No. 8422890) [Doctoral dissertation, University of Pennsylvania]. ProQuest Dissertations Publishing.

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, *1*, 1–47.

Carranza, I. (1995). Multilevel analysis of two-way immersion discourse. In J. Alatis, C. Straehle, B. Gallenberger, & M. Ronkin (Eds.), *Georgetown University round table on languages and linguistics* (pp. 169–187). Georgetown University Press.

Carroll, Susan. (1982). Redoublement et dislocations en français populaire. In C. Lefebvre (Ed.), *La syntaxe comparée du français standard et populaire: Approches formelles et fonctionnelles* (Vol. 1) (pp. 290–357). Gouvernement du Québec, Office de la Langue Française.

Carton, F. (1987). Les accents régionaux. In G. Vermès & J. Boutet (Eds.), *France pays multilingue, vol. 2 (Pratiques des langues)* (pp. 29–49). L'Harmattan.

Carton, F. (2009). Etude prosodique d'un cas de détachement : les pronoms personnels pseudo-disjoints dans un corpus de presse parlée en français. http://carton.fernand.free.fr/Pronoms%20Nancy%20revu.pdf

Castañeda, M.E., & Zirger, M.L. (2011). Making the most of the new study abroad: Social capital and the short-term sojourn. *Foreign Language Annals*, *44*, 544–64.

Chafe, W. (1994). *Discourse, Consciousness, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing*. University of Chicago Press.

Chanet, C. (2001). 1700 occurrences de la particule *quoi* en français parlé contemporain : approche de la « distribution » et des fonctions en discours. *Marges Linguistiques*, *2*, 56–80.

Chesterman, A. (1991). *On definiteness*. Cambridge University Press.

Chevalier, G. (2001). Comment *comme* fonctionne d'une génération à l'autre. *Revue Québécoise de Linguistique*, *30*(2), 13–40.

Chifflet, L. (1659). *Essay d'une parfaite grammaire de la langue françoise*. Slatkine Reprints, 1973.

Chomsky, N. (1995). *The minimalist program*. MIT Press.

Cinque, G. (1990). *Types of A-bar dependencies, Vol. 17 of Linguistic inquiry monographs*. MIT Press.

Clark, E.V., & de Marneffe, M.-C. (2012). Constructing verb paradigms in French: Adult construals and emerging grammatical contrasts. *Morphology*, *22*, 89–120.

Combaz, C. (2014, September 18). "'La France, elle est…' : Parlez-vous le François Hollande ?" *Le Figaro*. https://www.lefigaro.fr/vox/politique/2014/09/18/31001-20140918ARTFIG00383-la-france-elle-est-parlez-vous-le-francois-hollande.php

Comrie, B. (1981). *Language universals and linguistic typology*. University of Chicago Press.

Cook, V. (2005). Basing teaching on the L2 user. In E. Llurda (Ed.), *Non-native language teachers: Perceptions, challenges and contributions to the profession* (pp. 217–41). Springer.

Coppieters, R. (1987). Competence differences between native and near-native speakers. *Language*, *63*, 544–573.

Coveney, A. (1996/2002). *Variability in spoken French: A sociolinguistic study of interrogation and negation*. Elm Bank.

Coveney, A. (1998). Awareness of linguistic constraints on variable *ne* omission. *Journal of French Language Studies*, *8*(2), 159–187.

Coveney, A. (2003). Le redoublement du sujet en français parlé : une approche variationniste. *Etudes Romanes*, *54*, 111–143.

Coveney, A. (2005). Subject doubling in spoken French: A sociolinguistic approach. *The French Review*, *79*(1), 96–111.

Cubillos, J. H. (2013). Community engagement and proficiency gains in short-term study abroad programs. *NECTFL Review*, *71*, 17–36.

Cubillos, J. H., Chieffo, L., & Fan, C. (2008). The impact of short-term study abroad programs on listening comprehension skills. *Foreign Language Annals*, *41*, 157–86.

Culbertson, J. (2010). Convergent evidence for categorical change in French: From subject clitic to agreement marker. *Language*, *86*(1), 85–132.

Dannequin, C. (1977). *Les enfants baillonnés.* Nathan.

Davidson, D. E. (2007). Study abroad and outcome measurements: The case of Russian. *Modern Language Journal*, *91*, 276–80.

Davidson, D. E. (2010). Study abroad: When, how long and with what results? New data from the Russian front. *Foreign Language Annals*, *43*, 6–26.

De Cat, C. (2005). French subject clitics are not agreement markers. *Lingua*, *115*, 1195–1219.

De Cat, C. (2007). French dislocation without movement. *Natural Language and Linguistic Theory*, *25*, 485–534.

Dekydtspotter, L. (1993). The syntax and semantics of the French *ne que* construction. In U. Lahiri & A. Z. Wyner (Eds.), *Semantics and Linguistic Theory (SALT) 3* (pp. 38–56). CLC Publications.

Dekydtspotter, L., & Petrush, R.A. (2006). On the acquisition of *ne ... que* 'only' in English-French. In M. Grantham O'Brien, C. Shea, & J. Archibald (Eds.), *Proceedings of the 8th Generative Approaches to Second Language Acquisition Conference* (pp. 25–32). Cascadilla Proceedings Project.

Demuth, K., & Tremblay, A. (2008). Prosodically-conditioned variability in children's production of French determiners. *Journal of Child Language*, *35*, 99–127.

Deshaies, D., Guilbault, C., & Paradis, C. (1993). Prosodie et dislocation à gauche par anaphore en français québécois spontané. In A. Crochetière, J. C. Boulanger & C. Ouellon (Eds.), *Actes du XVe Congrès international des linguistes* (pp. 31–34). Les Presses de L'Université Laval.

Develey, A. (2017, April 30). "Paris est beau ou belle ?" *Le Figaro*. https://www.lefigaro.fr/langue-francaise/expressions-francaises/2017/04/30/37003-20170430ARTFIG00002-paris-est-beau-ou-belle.php

Dewaele, J.-M. (1992). L'omission du "ne" dans deux styles oraux d'interlangue française. *Journal of Applied Linguistics*, *7*(1), 3–17.

Dewaele, J.-M. (1999). Word order variation in interrogative structures of native and non-native French. *International Journal of Applied Linguistics*, *123*, 161–180.

Dewaele, J.-M. (2002). Using socio-stylistic variants in advanced French interlanguage: The case of 'nous' / 'on'. In S. Foster-Cohen, T. Ruthenberg, & M. L. Poschen (Eds.), *Eurosla Yearbook* (Vol. 2, pp. 205–226). John Benjamins.

Dewaele, J.-M. (2004a). Retention or omission of the *ne* in advanced French interlanguage: The variable effect of extralinguistic factors. *Journal of Sociolinguistics*, *8*, 433–450.

Dewaele, J.-M. (2004b). Vous or tu? Native and non-native speakers of French on a sociolinguistic tightrope. *International Review of Applied Linguistics*, *42*, 383–402.

Dewaele, J.-M. (2007). Diachronic and/or synchronic variation? The acquisition of sociolinguistic competence in L2 French. Birkbeck ePrints. https://eprints.bbk.ac.uk/id/eprint/716/

Dewaele, J.-M., & Regan, V. (2001). The use of colloquial words in advanced French interlanguage. In S. Foster-Cohen (Ed.), *EUROSLA Yearbook 2001* (pp. 51–68). John Benjamins.

Dewaele, J.-M., & Regan, V. (2002). Maîtriser la norme sociolinguistique en interlangue française: Le cas de l'omission variable de 'ne'. *Journal of French Language Studies*, *12*, 123–148.

Di Silvio, F., Donovan, A., & Malone, M. (2014). The effect of study abroad homestay placements: Participant perspectives and oral proficiency gains. *Foreign Language Annals*, *47*, 168–188.

Diller, A.-M. (1983). Subject NP structure and variable constraints: The case of *ne* deletion. In R. Fasold (Ed.), *Variation in the form and use of language: A sociolinguistics reader* (pp. 167–174). Washington, DC: Georgetown University Press.

Donaire, M. (2013). Entité lexicale : enfin. In J. Anscombre, M. Donaire, & P. Haillet (Eds.), *Opérateurs discursifs du français : Eléments de description sémantique et pragmatique* (pp. 105–122). Peter Lang.

Donaldson, B. (2008). *Discourse competence in near-native speakers of French* (Publication No. 3319913) [Doctoral dissertation, Indiana University Bloomington]. ProQuest Dissertations Publishing.

Donaldson, B. (2011a). Left dislocation in near-native French. *Studies in Second Language Acquisition*, *33*, 399–432.

Donaldson, B. (2011b). Nativelike right-dislocation in near-native French. *Second Language Research*, *27*, 361–390.

Donaldson, B. (2017). Negation in near-native French: Variation and sociolinguistic competence. *Language Learning*, *67*, 141–170.

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition.* Lawrence Erlbaum.

Dubuisson, C., Emirkanian, L., & Lemay, L. (1983). La dislocation, la topicalisation et la non-reprise du sujet sont-elles des critères de maturation syntaxique ? *Revue de l'Association Québécoise de Linguistique*, *2*(3), 23–33.

Dupont, N. (1985). *Linguistique du détachement en français.* Peter Lang.

Dye, C.D. (2011). Reduced auxiliaries in early child language: converging observational and experimental evidence from French. *Journal of Linguistics*, *47*, 301–339.

Ellis, R. (1985). *Understanding second language acquisition*. Oxford University Press.

Ellis, R. (1987). Interlanguage variability in narrative discourse: Style shifting in the use of past tense. *Studies in Second Language Acquisition*, *9*, 1–20.

Ellis, R. (1999). Item versus system learning: Explaining free variation. *Applied Linguistics*, *20*, 460–480.

Etienne, C., & Sax, K. (2009). Stylistic variation in French: Bridging the gap between research and textbooks. *The Modern Language Journal*, *93*, 584–606.

Fagyal, Z., Kibbee, D., & Jenkins, F. (2006). *French: A linguistic introduction.* Cambridge University Press.

Ferdinand, A. (2002). Acquisition of syntactic topic marking in L2 French. *Linguistics in the Netherlands 2002*, 49–59.

Ferguson, C. (1971). Absence of copula and the notion of simplicity: A study of normal speech, baby talk, foreigner talk and pidgins. In D. Hymes (Ed.), *Pidginization and creolization of languages* (pp. 141–150). Cambridge University Press.

Fernández Dobao, A. (2018). Collaborative dialogue in learner-learner and learner-native speaker interaction. *Applied Linguistics*, *33*(3), 229–256.

Ferrand, F. (2015, February 9). "Par pitié M. Hollande, épargnez la langue française." *Le Figaro*. https://www.lefigaro.fr/vox/societe/2015/02/09/31003-20150209ARTFIG00340-franck-ferrand-par-pitie-m-hollande-epargnez-la-langue-francaise.php

Fishman, J. (1965). Who speaks what language to whom and when? *La Linguistique*, *1*(2), 67–88.

Fleischman, S., & Yaguello, M. (2004). Discourse markers across languages? Evidence from English and French. In C. L. Moder & A. Martinovic-Zik (Eds.), *Discourse across Languages and Cultures* (pp. 129–147). John Benjamins.

Flikeid, K., & Péronnet, L. (1989). N'est-ce pas vrai qu'il faut dire 'J'avons été'?: Divergences régionales en acadien. *Français Moderne*, *57*, 219–228.

Fonseca-Greber, B. B. (2000). *The change from pronoun to clitic to prefix and the rise of null subjects in spoken Swiss French* (Publication No. 9992079) [Doctoral dissertation, University of Arizona]. ProQuest Dissertations Publishing.

Fonseca-Greber, B. B. (2007). The emergence of emphatic 'ne' in conversational Swiss French. *Journal of French Language Studies*, *17*, 219–275.

Fonseca-Greber, B. B., & L. Waugh. (2002). The subject clitics of European conversational French: Morphologization, grammatical change, semantic change, and change in progress. In R. Nùñez-Cedeño, L. López, & R. Cameron (Eds.), *Proceedings of the 31st Symposium in the Romance Languages* (pp. 99–117). John Benjamins.

Forsberg, F. F., Bartning, I., Engel, H., Gudmundson, A., Hancock, V., & Lindqvist, C. (2014). Beyond advanced stages in high-level spoken L2 French. *Journal of French Language Studies*, *24*, 255–280.

Freed, B.F. (1990). Language learning in a study abroad context: The effects of interactive and non-interactive out-of-class contact on grammatical achievement and oral proficiency. In J. Atlatis (Ed.), *Linguistics, language teaching and language acquisition: The interdependence of theory, practice and research* (Georgetown University Round Table on Languages and Linguistics) (pp. 459–477). Georgetown University Press.

Freed, B.F. (Ed.) (1995). *Second language acquisition in a study abroad context.* John Benjamins.

French, L., & Beaulieu, S. (2016). Effects of sociolinguistic awareness on French L2 learners' planned and unplanned oral production of stylistic variation. *Language Awareness*, *25*, 55–71.

Gaatone, D. (1999). Réflexions sur la syntaxe de 'ne...que'. In M. Plénat, M. Aurnague, A. Condamines, J.P. Maurel, C. Molinier, & C. Muller (Eds.), *L'emprise du sens: Structures linguistiques et interprétations* (pp. 101–115). Rodopi.

Gadet, F. (1992). *Le français populaire*. Presses Universitaires de France.

Gadet, F. (1997). *Le français ordinaire* (2nd edition). Colin.

Garrod, S., & Pickering, M. (2004). Why is conversation so easy? *Trends in Cognitive Sciences*, *8*(1), 8–11.

Gass, S. M., Behney, J., & Plonsky, L. (2013). *Second language acquisition: An introductory course*. Routledge.

Gass, S. M., Mackey, A., & Pica, T. (1998). The role of input and interaction in second language acquisition: Introduction to the Special Issue. *Modern Language Journal*, 82, 299–305.

Gaucher, D. (2013). *The agreement of the past participle in spoken French, as a sociolinguistic variable: Production and perception* (Publication No. U618805) [Doctoral Dissertation, University of Exeter]. ProQuest Dissertations Publishing.

Geeslin, K. (2011). The acquisition of variation in second language Spanish: How to identify and catch a moving target. In M. Díaz-Campos (Ed.), *The Handbook of Hispanic sociolinguistics* (pp. 303–320). Wiley-Blackwell.

Geeslin, K. (2019, September). *Variationist sociolinguistics and SLA: Current insights and future directions* [Conference paper]. Second Language Research Forum, Michigan State University.

Geeslin, K. (2020). Variationis perspective(s) on interlocutor individual differences. In L. Gurzynski-Weiss (Ed.), *Cross-theoretical explorations of interlocutors and their individual differences* (pp. 129–159). John Benjamins.

Geeslin, K., Fafulas, S., & Kanwit, M. (2013). Acquiring geographically-variable norms of use: The case of the present perfect in Mexico and Spain. In C. Howe, S. E. Blackwell, & M. Lubbers Quesada (Eds.), *Selected proceedings of the 15th Hispanic Linguistics Symposium* (pp. 205–220). Cascadilla Proceedings Project.

Geeslin, K., Garcia-Amaya, L., Hasler, M., Henrikson, N., & Killam, J. (2010). The SLA of direct object pronouns in an immersion environment where use is variable. In C. Borgonovo, M. Español-Echevarría & P. Prevost (Eds.), *Refereed Proceedings from the 12th Hispanic Linguistics Symposium* (pp. 246–259). Cascadilla Press.

Geeslin, K., Garcia-Amaya, L., Hasler, M., Henrikson, N., & Killam, J. (2012). The L2 acquisition of variable perfective past time reference in Spanish in an overseas immersion setting. In K. Geeslin & M. Díaz-Campos (Eds.), *Refereed Proceedings for the 14th Hispanic Linguistics Symposium* (pp. 197–213). Cascadilla Press.

Geeslin, K. L. & Gudmestad, A. (2011). Using sociolinguistic analyses of discourse-level features to expand research on L2 variation in forms of Spanish subject expression. In L. Plonsky & M. Schierloh (Eds.), *Selected proceedings of the 2009 Second Language Research Forum: Diverse contributions to SLA* (pp. 16–30). Cascadilla Press.

Geeslin, K., & Long, A. (2014). *Sociolinguistics and second language acquisition: Learning to use language in context.* Routledge.

Geluykens, R. (1992). *From discourse process to grammatical construction: On left dislocation in English*. Benjamins.

George, A. (2013). The development of /θ/, a variable geographic phonetic feature, during a semester abroad: The role of explicit instruction. In J. Levis & K. LeVelle (Eds.),

Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference (pp. 120–128). Iowa State University.

George, A. (2014). Study abroad in Central Spain: The development of regional phonological features. *Foreign Language Annals*, *47*, 97–114.

George, A. (2018). The development of a regional morphosyntactic feature by learners of Spanish in a study abroad setting: The case of *vosotros*. *Hispanic Studies Review*, *3*, 101–125.

George, K. (1993). Alternative French. In C. Sanders (Ed.), *French today: Language in its social context* (pp. 155–170). Cambridge University Press.

Giles, H., & Powesland, P. (1975). *Speech style and social evaluation*. Academic Press London.

Giles, H., & Smith, P. (1979). Accommodation Theory: Optimal levels of convergence. In H. Giles & R. St. Clair (Eds.), *Language and social psychology* (pp. 45–65). Basil Blackwell.

Glahn, E. (1993). An analysis of learner-native speaker dyadic interaction with regard to dynamics, coherence, and dominance. In B. Kettemann & W. Wieden (Eds.), *Current issues in European second language acquisition research* (pp. 327–335). Gunther Narr.

Gnevsheva, K. (2017). Within-speaker variation in passing for a native speaker. *International Journal of Bilingualism*, *21*(2), 213–227.

Gotowski, M. (2015). Subject doubling in child French. *University of Pennsylvania Working Papers in Linguistics*, *21*(1), Article 12.

Gougenheim, G., Rivenc, P., Michéa, R., & Sauvageot, A. (1964). *L'élaboration du français fondamental*. Didier.

Gregory, M. L., & Michaelis, L. A. (2001). Topicalization and left dislocation: A functional opposition revisited. *Journal of Pragmatics*, *33*, 1665–1706.

Gudmestad, A. (2014). Variationist approaches to second language Spanish. In K. L. Geeslin (Ed.), *Handbook of Spanish second language acquisition* (pp. 80–95). Wiley-Blackwell.

Guilbault, C. (1995). *La dislocation à gauche en français québécois spontané: Analyse instrumentale* (Publication No. MM02374) [Master's Thesis, Université Laval]. ProQuest Dissertations Publishing.

Guillon, J.-C. (2002). La norme, l'usage et le maître (4e partie): bis repetita ou le redoublement du sujet. *Connaissance du français*, *45*, 35–46.

Gurzynski-Weiss, L. (2010). *Factors influencing oral corrective feedback provision in the Spanish foreign language classroom: Investigating instructor native/nonnative speaker status, SLA education, & teaching experience* (Publication No. AAI3412596) [Doctoral dissertation, Georgetown University]. ProQuest Dissertations Publishing.

Gurzynski-Weiss, L. (2013). Instructor characteristics and classroom-based second language acquisition of Spanish. In K. Geeslin (Ed.), *The handbook of Spanish second language acquisition* (pp. 530–546). Wiley-Blackwell.

Gurzynski-Weiss, L., & Plonsky, L. (2017). Look who's interacting: A scoping review of research involving non-teacher/non-peer interlocutors. In L. Gurzynski-Weiss (Ed.), *Expanding individual difference research in the interaction approach: Investigating learners, instructors, and other interlocutors* (pp. 305–324). John Benjamins.

Haileselassie, A. (2015). *Voilà, an orientation shift marker in modern French discourse: A conversation analytic perspective.* (Publication No. 3740460) [Doctoral dissertation, University of Illinois at Urbana-Champaign]. ProQuest Dissertations Publishing.

Hansen, A. B., & Malderez, I. (2004). Le *ne* de négation en région parisienne: Une étude en temps réel. *Langage et société*, *107*, 5–30.

Hardison, D. (2014). Changes in second-language learners' oral skills and socio-affective profiles following study abroad: A mixed-methods approach. *The Canadian Modern Language Review*, *70*(4), 415–444.

Hendricks, H. (2000). The acquisition of topic marking in L1 Chinese and L1 and L2 French. *Studies in Second Language Acquisition*, *22*, 369–397.

Hennecke, I. (2017). The impact of pragmatic markers and hedging on sentence comprehension: A case study of *comme* and *genre*. *Journal of French Language Studies*, *27*, 355–380.

Hernández, T. (2016). Short-term study abroad: Perspectives on speaking gains and language contact. *Applied Language Learning*, *26*(1), 39–64.

Hirschbühler, P., & Labelle, M. (1994). L'évolution des propositions infinitives négatives en français. *Linx*, *5*, 59–90. https://doi.org/10.4000/linx.1194

Holmberg, A., & Nikanne, U. (2008). Subject doubling in Finnish: The role of deficient pronouns. In S. Barbiers, M. Lekakou, M. Van der Ham, & O. Koeneman (Eds.), *Microvariation in syntactic doubling* (pp. 325–349). Emerald.

Housen, A., Kuiken, K., & Vedder, I. (2012). *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*. John Benjamins.

Howard, M. (2012). The advanced learner's sociolinguistic profile: On issues of individual differences, second language exposure conditions, and type of sociolinguistic variable. *The Modern Language Journal*, *96*, 20–33.

Hummel, K. M. (2014). *Introducing second language acquisition: Perspectives and practices*. Wiley Blackwell.

Hunt, K. W. (1965). A synopsis of clause-to-sentence length factors. *English Journal*, *54*, 300–309.

Ioup, G., Boustagui, E., El Tigi, M., & Moselle, M. (1994). Reexamining the critical period hypothesis: A case study of successful adult SLA in a naturalistic environment. *Studies in Second Language Acquisition*, *16*, 73–98.

Isambert, P. (2016). *Genre*: un mode récente mais qui vient de loin. *Journal of French Language Studies*, *26*, 85–96.

Itakura, H. (2001). Describing conversational dominance. *Journal of Pragmatics*, *33*, 1859–80.

Jandry Hertel, T., & Sunderman, G. (2009). Student attitudes toward native and non-native language instructors. *Foreign Language Annals*, *42*(3), 468–482.

Jespersen, O. (1917). *Negation in English and other languages*. Høst.

Johnson, D. (2009). Getting off the Goldvarb standard: introducing Rbrul for mixed-effects variable rule analysis. *Language and Linguistics Compass*, *3*(1), 359–83.

Kanwit, M. (2018). Variation in second-language Spanish. In K. L. Geeslin (Ed.), *The Cambridge Handbook of Spanish Linguistics,* (pp. 716–736). Cambridge University Press.

Kanwit, M., Elias, V., & Clay, R. (2018). Acquiring intensifier variation abroad: Exploring *muy* and *bien* in Spain and Mexico. *Foreign Language Annals*, *51*, 455–471.

Kanwit, M., Geeslin, K., & Fafulas, S. (2015). The role of geography in the SLA of variable structures: A look at the present perfect, the copula contrast, and the present progressive in Mexico and Spain. *Probus*, *27*(2), 307–348.

Kanwit, M., & Solon, M. (2013). Acquiring variation in future-time expression abroad in Valencia, Spain and Mérida, Mexico. In J. Cabrelli Amaro, G. Lord, A. de Prada Pérez, &

J. E. Aaron (Eds.), *Selected proceedings of the 16th Hispanic Linguistics Symposium* (pp. 206–221). Cascadilla Proceedings Project.

Kasper-Cushman, K. (forthcoming). *Mapping the functional distribution of human impersonal uses of on, ça, and ils, in French*. [Doctoral Dissertation, Indiana University Bloomington].

Kayne, R. (1975). *French syntax: The transformational cycle*. MIT Press.

Kayne, R. (1991). Romance clitics, verb movement, and PRO. *Linguistic Inquiry*, *22*, 647–686.

Kayne, R. (1994). *The antisymmetry of syntax*. MIT Press.

Kennedy, K. (2012). *What we don't learn in the classroom: The acquisition of sociolinguistic competence during study abroad* (Publication No. AAI 3555343) [Doctoral dissertation, University of California Davis]. ProQuest Dissertations Publishing.

Kennedy Terry, K. (2017). Contact, context, and collocation: The emergence of sociostylistic variation in L2 French learners during study abroad. *Studies in Second Language Acquisition*, *39*, 553–578.

Kerr, B. J. (2002). Variant word-order constructions: To teach or not to teach—Evidence from learner narratives. In S. M. Gass, K. Bardovi-Harlig, S. S. Magnan, & J. Walz (Eds.), *Pedagogical norms for second and foreign language learning and teaching* (pp. 183–198). John Benjamins.

King, R., & Nadasdi, T. (1997). Left dislocation, number marking, and (non-)standard French. *Probus*, *9*, 267–284.

Kinginger, C., & Carnine, J. (2019). Language learning at the dinner table: Two case studies of French homestays. *Foreign Language Annals*, *52*, 850–872.

Knouse, S. M. (2012). The acquisition of dialectal phonemes in a study abroad context: The case of the Castilian theta. *Foreign Language Annals*, *45*, 512–542.

Kohn, K. (1982). Beyond output: the analysis of interlanguage development. *Studies in Second Language Acquisition*, *4*(2), 137–152.

Kramsch, C. (2002). Standard, norm, and variability in language learning: A view from foreign language research. In S. M. Gass, K. Bardovi-Harlig, S. Sieloff Magnan, & J. Walz (Eds.), *Pedagogical norms for second and foreign language learning and teaching* (pp. 59–80). John Benjamins.

Labelle, G. (1976). La langue des enfants de Montréal et de Paris. *Langue française*, *31*, 55–73.

Labelle, M. (2019). The French Jespersen's cycle and negative concord. In D. L. Arteaga (Ed.), *Contributions of Romance languages to current linguistic theory* (*Studies in natural language and linguistic theory, vol. 95*) (pp. 155–172). Springer.

Labov, W. (1972). *Sociolinguistic patterns.* University of Pennsylvania Press.

Laenzlinger, C. (1998). *Comparative studies in word order variation*. John Benjamins.

Lafontaine, D. (1986). *Le parti pris des mots. Normes et attitudes linguistiques.* Mardaga.

Lambrecht, K. (1981). *Topic, antitopic, and verb agreement in non-standard French*. John Benjamins.

Lambrecht, K. (1987). On the status of SVO sentences in French discourse. In R. Tomlin (Ed.), *Coherence and grounding in discourse* (pp. 217–261). John Benjamins.

Lambrecht, K. (1994). *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents*. John Benjamins.

Lambrecht, K. (2001). Dislocation. In M. Haspelmath, E. König, W. Oesterreicher, & W. Raible (Eds.), *Language typology and language universals, vol. 2* (pp. 1050–1078). Walter de Gruyter.

Lardiere, D. (2007). *Ultimate attainment in second language acquisition: A case study.* Lawrence Erlbaum.

Larrivée, P. (2014). Reanalysis of negatives as polarity markers? The last 400 years of decline of the French preverbal negative clitic. *Lingua*, *147*, 40–49.

Larsen-Freeman, D. E. (1975). The acquisition of grammatical morphemes by adult ESL students. *TESOL Quarterly*, *9*, 409–419.

Lasagabaster, D., & Sierra, J. M. (2005). What do students think about having a NS teacher? In E. Llurda (Ed.), *Non-native language teachers: Perceptions, challenges and contributions to the profession* (pp. 217–41). Springer.

Laurendeau, P., Néron, M., & Fournier, R. (1982). Contraintes sur l'emploi du pro-écho sujet en français du Québec. *Revue de l'Association Québécoise de Linguistique*, 115–128.

Leal Méndez, T., Rothman, J., & Slabakova, R. (2015). Discourse-sensitive clitic-doubled dislocations in heritage Spanish. *Lingua*, *155*, 85–97.

Lee, H., Joo, K., Moon, J., & Hong, Y. (2006). The effects of interlocutor type on learners' output production: NS teacher, NNS teacher and NNS peers. *Korean Journal of Applied Linguistics*, *22*(2), 201–220.

Lemieux, M. (1985). Pas rien. In M. Lemieux & H. Cedergren (Eds.), *Les tendances dynamiques du français parlé à Montréal, tomes 1 et 2*. Office de la langue française.

Lennon, P. (1993). The advanced learner: Affective, social and motivational factors. *The Language Learning Journal*, *8*(1), 39–43.

Leow, R. (2001). Attention, awareness, and foreign language behavior. In R. Ellis (Ed.), *Form-focused instruction and second language learning* (pp. 113–155). Blackwell Publishing.

Li, X. (2017). Stylistic variation in L1 and L2 Chinese: Native speakers, learners, teachers and textbooks. *Chinese as a Second Language*, *52*(1), 55–77.

Linford, B. G. (2016). *The second-language development of dialect-specific morpho-syntactic variation in Spanish during study abroad* (Publication No. 10130845) [Doctoral dissertation, Indiana University Bloomington]. ProQuest Dissertations Publishing.

Loban, W. (1976). *Language development: Kindergarten through grade twelve*. National Council of Teachers of English.

Lodge, A. (2007). The history of French. In Ayoun, D. (Ed.), *French applied linguistics* (pp. 21–48). John Benjamins.

Long, M. (1981). Input, interaction, and second language acquisition. In H. Winitz (Ed.), *Native languages and foreign language acquisition, annals of the New York Academy of Sciences, 379* (pp. 259–278). New York Academy of Sciences.

Long, M. (1990). The least a Second Language Acquisition theory needs to explain. *TESOL Quarterly*, 24, 649–666.

Lüdicke, A. (1982). Zum Ausfall der Verneinungspartikel *ne* im gesprochenen Französisch. *Zeitschrift für Romanische Philologie*, *98*, 43–58.

Lyster, R. (1993). *The effect of functional-analytic teaching on aspects of sociolinguistic competence: A study in French immersion classrooms at the grade eight level* (Publication No. NN82719) [Doctoral dissertation, University of Toronto]. ProQuest Dissertations Publishing.

Mackey, A. (Ed.) (2012). *Input, interaction, and corrective feedback in L2 classrooms*. Oxford University Press.

Mackey, A., Oliver, R., & Leeman, J. (2003). Interactional input and the incorporation of feedback: An exploration of NS-NNS and NNS-NNS adult and child dyads. *Language Learning*, *53*(1), 35–66.

Magnan, S.S., & Back, M. (2007). Social interaction and linguistic gain during study abroad. *Foreign Language Annals*, *40*, 43–61.

Marceau, J. (1977). *Class and status in France: economic change and social immobility, 1945-1975.* Oxford University Press.

Martineau, F., & Mougeon, R. (2003). A sociolinguistic study of the origins of *ne* deletion in European and Quebec French. *Language*, *79*(1), 118–152.

Martinsen, R. (2010). Short-term study abroad: Predicting changes in oral skills. *Foreign Language Annals*, *43*, 504–530.

Massot, B. (2010). Le patron diglossique de variation grammaticale en français. *Langue Française*, *168*, 87–106.

Matthews, P. (1997). *The concise Oxford dictionary of linguistics*. Oxford University Press.

Meisner, C. (2016). *La variation pluridimensionnelle : Une analyse de la négation en français*. Peter Lang.

Meisner, C., & Pomino, N. (2014). Synchronic variation in the expression of French negation: A Distributed Morphology approach. *Journal of French Language Studies*, *24*, 9–28.

Mihatsch, W. (2006). *Machin, truc, chose* : la naissance de marqueurs pragmatiques. In M. Drescher & B. Job (Eds.), *Les marqueurs discursifs dans les langues romanes : approches théoriques et méthodologiques* (pp. 153–172). Peter Lang.

Montrul, S. (2011). Multiple interfaces and incomplete acquisition. *Lingua*, *121*, 591–604.

Moreau, M.-L. (1986). Les séquences préformées : Entre les combinaisons libres et les idiomatismes. Le cas de la négation avec ou sans *ne*. *Le Français Moderne*, *54*, 137–160.

Morin, Y.-C. (1979). La morphophonologie des pronoms clitiques en français populaire. *Cahier de linguistique*, *9*, 1–36.

Morin, Y.-C. (1982). De quelques [l] non étymologiques dans le français du Québec : Notes sur les clitiques et la liaison. *Revue Québécoise de Linguistique*, *11*(2), 9–47.

Mougeon, R., & Beniak, E. (1991). *Linguistic consequences of language contact and restriction: The case of French in Ontario, Canada*. Oxford University Press.

Mougeon, R., & Rehner, K. (2017). The influence of classroom input and community exposure on the learning of variable grammar. *Bilingualism: Language and Cognition*, *20*, 21–22.

Mougeon, R., Nadasdi, T., & Rehner, K. (2002). État de la recherche sur l'appropriation de la variation par les apprenants avancés du FL2 ou FLE. *Acquisition et Interaction en Langue Etrangère*, *17*, 7–50.

Mougeon, R., Nadasdi, T., & Rehner, K. (2010). *The sociolinguistic competence of immersion students*. Multilingual Matters.

Moufflet, A. (1935). *Encore le massacre de la langue française*. Privat/Didier.

Nadasdi, T. (1995a). Subject NP doubling, matching, and minority French. *Language Variation and Change*, *7*, 1–14.

Nadasdi, T. (1995b). *Variation morphosyntaxique et langue minoritaire: le cas du français ontarien.* (Publication No. NN02820) [Doctoral dissertation, University of Toronto]. ProQuest Dissertations Publishing.

Nadasdi, T., Mougeon, R., & Rehner, K. (2008). Factors driving lexical variation in L2 French: A variationist study of automobile, auto, voiture, char and machine. *Journal of French Language Studies*, *18*, 365–381.

Nagy, N., Blondeau, H., & Auger, J. (2003). Second language acquisition and "real" French: An investigation of subject doubling in the French of Montreal Anglophones. *Language Variation and Change*, *15*, 73–103.

Nardy, A., Chevrot, J.-P., & Barbu, S. (2014). Sociolinguistic convergence and social interactions within a group of preschoolers: A longitudinal study. *Language Variation and Change*, *26*, 273–301.

Oliver, R. (2002). The patterns of negotiation for meaning in child interactions. *The Modern Language Journal*, *86*(1), 97–111.

Orton, J. (2014). Comparing teachers' judgments of learners' speech in Chinese as a Foreign Language. *Foreign Language Annals*, *47*(3), 507–526.

Ossipov, H. (1990). *A GPSG analysis of doubling and dislocation constructions in French* (Publication No. 9107285) [Doctoral dissertation, Indiana University Bloomington]. ProQuest Dissertations Publishing.

Ossipov, H. (2002). Dislocated subjects in French: A pedagogical norm. In S. M. Gass, K. Bardovi-Harlig, & S. S. Magan (Eds.), *Pedagogical norms for second and foreign language learning and teaching: Studies in honor of Albert Valdman* (pp. 171–181). John Benjamins.

Oudin, A. (1632). *Grammaire françoise*. Slatkine reprints, 1972.

Overstreet, M. (2005). And stuff *und so*: Investigating pragmatic expressions in English and German. *Journal of Pragmatics*, *37*(11), 1845–1864.

Painchaud, G., d'Anglejan, A., & Vincent, D. (1982). Acquisition du français par un groupe d'immigrants asiatiques. Research report, Faculté des Sciences de l'Education, Université de Montréal.

Palasis, K. (2010). Introducing new French child data: Thoughts on their gathering and coding. *Corpus*, *9*, 33–51.

Palasis, K. (2015). Subject clitics and preverbal negation in European French: Variation, acquisition, diatopy and diachrony. *Lingua*, *161*, 125–143.

Paradis, C. (1985). *An acoustic study of variation and change in the vowel system of Chicoutimi and Jonquière (Québec)* (Publication No. 8603685) [Doctoral dissertation, University of Pennsylvania]. ProQuest Dissertations Publishing.

Perdue, C. (Ed.). (1984). *Second language acquisition by adult immigrants: A field manual.* Newbury House.

Perdue, C., Deulofeu, J., & Trévise, A. (1992). The acquisition of French. In W. Klein & C. Perdue (Eds.), *Utterance structure: Developing grammars again* (pp. 225–300). John Benjamins.

Piller, I. (2002). Passing for a native speaker: Identity and success in second language learning. *Journal of Sociolinguistics*, *6*, 179–206.

Piron, S. (2017). *Grammaire française : perfectionnement vol. 2*. De Boeck Supérieur.

Pohl, J. (1968). Ne dans le français parlé contemporain: Les modalités de son abandon. In A. Quilis (Ed.), *Actas XI Congreso Internacional de Linguistica y Filologia Románicas, vol. 3* (pp. 1343–1359). Consejo Superior de Investigaciones Científicas.

Pohl, J. (1975). L'omission de *ne* dans le français contemporain. *Le français dans le Monde*, *117*, 17–23.

Pooley, T. (1996). *Chtimi: The urban vernaculars of northern France.* Multilingual Matters.

Poplack, S., & St-Amand, A. (2007). A real-time window on 19th-century vernacular French: The Récits du français québecois d'autrefois. *Language in Society*, *36*, 707–734.

Porter, P.A. (1983). *Variations in the conversations of adult learners of English as a function of the proficiency level of the participants* (Publication No. 8320763) [Doctoral dissertation, Stanford University]. ProQuest Dissertations Publishing.

Porter, P.A. (1986). How learners talk to each other: Input and interaction in task-centered discussions. In R. Day (Ed.), *Talking to learn: Conversation in second language acquisition* (pp. 200–222). Newbury House.

Posner, R. (1985). Post-verbal negation in non-standard French: A historical and comparative view. *Romance Philology*, *39*, 170–197.

Potowski, K. (2004). Student Spanish use and investment in a dual immersion classroom: Implications for second language acquisition and heritage language maintenance. *The Modern Language Journal*, *88*(1), 75–101.

Preston, D. R. (1991). Sorting out the variables in sociolinguistic theory. *American Speech*, *66*, 33–56.

Preston, D. R. (1996). Variationist perspectives on second language acquisition. In R. Bayley & D. R. Preston (Eds.), *Second language acquisition and linguistic variation* (pp. 1–45). John Benjamins.

Preston, D. R. (2004). Three kinds of sociolinguistics: A psycholinguistic perspective. In C. Fought (Ed.), *Sociolinguistic variation: Critical reflections* (pp. 140–158). Oxford University Press.

Priestley, L. (1955). Reprise constructions in French. *Archivum Linguisticum*, *7*(1), 1–28.

Pryde, M. (2014). Conversational patterns of homestay hosts and study abroad students. *Foreign Language Annals*, *47*(3), 487–506.

Quereuil, M. (1988). *La Bible française du XIIIe siècle. Édition critique de la Genèse*. Droz.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1972). *A grammar of contemporary English.* Longman Group.

Reaves, A. (2020). *Discourse markers in second language French* (Publication No. 28089381) [Doctoral dissertation, Indiana University Bloomington]. ProQuest Dissertations Publishing.

Regan, V. (1996). Variation in French interlanguage: A longitudinal study of sociolinguistic competence. In R. Bayley & D. Preston (Eds.), *Second language acquisition and linguistic variation* (pp. 177–201). John Benjamins.

Regan, V., Howard, M., & Lemée, I. (2009). *Acquisition of sociolinguistic competence in a study abroad context.* Multilingual Matters.

Rehner, K. (2002). *The development of aspects of linguistic and discourse competence by advanced second language learners of French* (Publication No. NQ74623) [Doctoral dissertation, University of Toronto]. ProQuest Dissertations Publishing.

Rehner, K. (2005). *Developing aspects of second language discourse competence*. Lincom Europa.

Rehner, K., & Mougeon, R. (1999). Variation in the spoken French of immersion students: To ne or not to ne, that is the sociolinguistic question. *Canadian Modern Language Review*, *56*, 124–154.

Rehner, K., Mougeon, R., & Nadasdi, T. (2003). The learning of sociolinguistic variation by advanced FSL learners. *Studies in Second Language Acquisition*, *25*, 127–156.

Renaud, C. (2010). *On the nature of agreement in English-French acquisition: A processing investigation in the verbal and nominal domains* (Publication No. AAI3423648) [Doctoral dissertation, Indiana University Bloomington]. ProQuest Dissertations Publishing.

Reynolds-Case, A. (2013). The value of short-term study abroad: An increase in students' cultural and pragmatic competency. *Foreign Language Annals*, *46*(2), 311–322.

Rickford, J., & McNair-Knox, F. (1994). Addressee- and topic-influenced style shift: A quantitative sociolinguistic study. In D. Biber & E. Finegan (Eds.), *Sociolinguistic perspectives on register* (pp. 235–276). Oxford University Press.

Ringer-Hilfinger, K. (2012). Learner acquisition of dialect variation in a study abroad context: The case of the Spanish [θ]. *Foreign Language Annals*, *45*, 430–446.

Rivers, W. P. (1998). Is being there enough? The effects of homestay placements on language gain during study abroad. *Foreign Language Annals*, *31*, 492–500.

Rizzi, L. (1986). On the status of subject clitics in Romance. In O. Jaeggli & C. Silva-Corvalàn (Eds.), *Studies in Romance linguistics* (pp. 391–419). Foris.

Roberge, Y. (1990). *The syntactic recoverability of null arguments*. McGill-Queen's University Press.

Rodrigues, I., & Neather, T. (2007). *Streetwise French*. Passport Books.

Rottet, K. (1995). *Language shift and language death in the Cajun French-speaking communities of Terrebonne and Lafourche parishes, Louisiana* (Publication No. 9539966) [Doctoral dissertation, Indiana University Bloomington]. ProQuest Dissertations Publishing.

Rowlett, P. (2013). Do French speakers really have two grammars? *Journal of French Language Studies*, *23*, 37–57.

Ryan, E. B., & Giles, H. (1982). *Attitudes towards language variation: Social and applied contexts*. Arnold.

Salgado-Robles, F. (2011). *The acquisition of sociolinguistic variation by learners of Spanish in a study abroad context* (Publication No. AAI3496927) [Doctoral dissertation, University of Florida Gainesville]. ProQuest Dissertations Publishing.

Sanell, A. (2007). *Parcours acquisitionnel de la négation et de quelques particules de portée en français L2.* University of Stockholm Press.

Sankoff, G. (1982). Usage linguistique et grammaticalisation : Les clitiques sujets en français. In N. Dittmar & B. Schlieben-Lange (Eds.), *La sociolinguistique dans les pays de langue romane* (pp. 81–85). Gunter Narr.

Sankoff, G., & Cedergren, H. (1972). Sociolinguistic research on French in Montréal. *Language in Society*, *1*(1), 173–74.

Sankoff, D., Tagliamonte, S., & Smith, E. (2005). Goldvarb X: A variable rule application for Macintosh and Windows. Department of Linguistics, University of Toronto.

Sankoff, G., Thibault, P., Nagy, N., Blondeau, H., Fonollosa, M.-O., & Gagnon, L. (1997). Variation in the use of discourse markers in a language contact situation. *Language Variation and Change*, *9*, 191–218.

Sankoff, G., & Vincent, D. (1977). L'emploi productif du *ne* dans le français parlé à Montréal. *Le français moderne*, *45*, 243–256.

Sax, K. J. (2003). *Acquisition of stylistic variation in American learners of French* (Publication No. 3122735) [Doctoral dissertation, Indiana University Bloomington]. ProQuest Dissertations Publishing.

Scarcella, R.C., & Higa, C.A. (1982). Input and age differences in second language acquisition. In S. Krashen, R.C. Scarcella, & M.H. Long (Eds.), *Child-adult differences in second language acquisition* (pp. 175–201). Newbury House.

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*, 129–158.

Schmidt-Rinehart, B. C., & Knight, S. M. (2004). The homestay component of study abroad: Three perspectives. *Foreign Language Annals*, *37*, 254–262.

Schumann, J. (1978). *The pidginization process: A model for second language acquisition*. Newbury House.

Segalowitz, N. (1976). Communicative incompetence and the non-fluent bilingual. *Canadian Journal of Behavioural Science*, *8*, 121–131.

Segalowitz, N., & Freed, B. (2004). Context, contact, and cognition in oral fluency acquisition: Learning Spanish in at home and study abroad contexts. *Studies in Second Language Acquisition*, *26*, 173–199.

Slabakova, R. (2006). Is there a critical period for semantics? *Second Language Research*, *22*(3), 302–338.

Slabakova, R. (2014). The bottleneck of second language acquisition. *Foreign Language Teaching and Research*, *46*(4), 543–559.

Slabakova, R. (2019). The Bottleneck Hypothesis updated. In T. Ionin & M. Rispoli (Eds.), *Three streams of generative language acquisition research: Selected papers from the 7th*

meeting of Generative Approaches to Language Acquisition – North America (GALANA-7) (pp. 319–345). John Benjamins.

Slabakova, R., Kempchinsky, P., & Rothman, J. (2012). Clitic-doubled left dislocation and focus fronting in L2 Spanish: A case of successful acquisition at the syntax-discourse interface. *Second Language Research*, *28*(3), 319–343.

Sleeman, P. (2004). Guided learners of French and the acquisition of emphatic constructions. *International Review of Applied Linguistics*, *42*, 129–151.

Sorace, A. (2006). Gradedness and optionality in mature and developing grammars. In G. Fanselow, C. Fery, M. Schlesewsky, & R. Vogel (Eds.), *Gradience in grammar: Generative perspectives* (pp. 106–123). Oxford University Press.

Sorace, A. (2011). Pinning down the concept of "interface" in bilingualism. *Linguistic Approaches to Bilingualism*, *1*, 1–33.

Sorace, A., & Serratrice, L. (2009). Internal and external interfaces in bilingual language development: Beyond structural overlap. *International Journal of Bilingualism*, *13*, 1–16.

Southard, B., & Muller, A. (1998). Blame it on Twain: Reading American dialects in *The Adventures of Huckleberry Finn*. In D. Oaks (Ed.), *Linguistics at work: A reader of applications* (pp. 565–573). Harcourt Brace.

Spithoff, M. (2018). *La chute et le maintien du ne de la négation en français parlé contemporain* [Master's thesis, Leiden University]. https://openaccess.leidenuniv.nl/handle/1887/65371

Stark, E. (1999). Antéposition et marquage du thème (topique) dans les dialogues spontanés. In C. Guimier (Ed.), *La thématisation dans les langues. Actes du colloque de Caen, 9-11 octobre 1997. Sciences pour la communication 53* (pp. 337–358). Lang.

Stark, E. (2012). Negation marking in French text messages. *Lingvisticae Investigationes*, *35*, 341–366.

Stark, E. (2013). Clitic subjects in French text messages: Does technical change provoke and/or reveal linguistic change? In K. J. Kragh & J. Lindschouw (Eds.), *Deixis and pronouns in Romance languages* (pp. 147–169). John Benjamins.

Sturm, J. (1981). *Morpho-syntaktische Untersuchungen zur phrase négative im gesprochenen Französisch*. *Die Negation mit und ohne NE*. Lang.

Swacker, M. E. (1978). *Attitudes of native and nonnative speakers toward varieties of American English* (Publication No. 7806839) [Doctoral dissertation, Texas A&M University]. ProQuest Dissertations Publishing.

Tagliamonte, S., & Jankowski, B. (2019). Grammatical convergence or microvariation? Subject doubling in English in a French dominant town. *Proceedings of the Linguistic Society of America*, *4*(17), 1–15. https://doi.org/10.3765/plsa.v4i1.4514

Tamminga, M. (2016). Persistence in phonological and morphological variation. *Language Variation and Change*, *28*, 335–356.

Tarone, E. (1979). Interlanguage as chameleon. *Language Learning*, *29*, 181–191.

Tarone, E. (1985). Variability in interlanguage use: A study of style-shifting in morphology and syntax. *Language Learning*, *35*, 373–404.

Thakerer, J., Giles, H., & Chesire, J. (1982). Psychological and linguistic parameters of speech accommodation theory. In C. Fraser & K. Scherer (Eds.), *Advances in the social psychology of language* (pp. 205–255). Cambridge University Press.

Thibault, P. (1983). *Equivalence et grammaticalisation* [Doctoral dissertation, Université de Montréal].

Thibault, P., & Sankoff, G. (1997). *L'insertion d'une langue seconde dans le répertoire d'une communauté linguistique: Le français parlé par les Anglo-Montréalais* [Conference paper]. NWAV 26, Laval University, Quebec City.

Thomas, A. (2004). Phonetic norm versus usage in advanced French as a second language. *International Review of Applied Linguistics in Language Teaching (IRAL)*, *42*, 365–382.

Torreira, F., Adda-Decker, M., & Ernestus, M. (2010). The Nijmegen Corpus of Casual French. *Speech Communication*, *52*, 201–212.

Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics*, *17*(1), 84–115.

Trévise, A. (1986). Is it transferable, topicalization? In E. Kellerman & M. Sharwood Smith (Eds.), *Crosslinguistic influence in second language acquisition* (pp. 186–206). Pergamon.

Trévise, A., & Noyau, C. (1984). Adult Spanish speakers and the acquisition of French negation forms: Individual variation and linguistic awareness. In R. Anderson (Ed.), *Second languages: A cross-linguistic perspective* (pp. 165–189). Newbury House.

Truby, K. (1995). *Deixis in discourse: the case of "ça, là" and "moi" in modern spoken French* (Publication No. 9543940) [Doctoral dissertation, University of California Davis]. ProQuest Dissertations Publishing.

Tsimpli, I., & Sorace, A. (2006). Differentiating interfaces: L2 performance in syntax-semantics and syntax-discourse phenomena. In D. Bamman, T. Magnitskaia, & C. Zaller (Eds.), *Proceedings of the 30th annual Boston University Conference on Language Development* (pp. 653–664). Cascadilla Press.

Tyne, H. (2004). Le style en français langue seconde: Vers un regard sociolinguistique. In F. Gadet (Ed.), *Le style comme perspective sur la dynamique des langues*. Special issue of *Langage et Société*, *109*, 31–51.

Uritescu, D., Mougeon, R., & Handouleh, Y. (2002). Le comportement du schwa dans le français parlé par les élèves des programmes d'immersion française. In C. Tatilon & A. Baudot (Eds.), *La linguistique fonctionnelle au tournant du siècle: Actes du 24e colloque international de linguistique fonctionnelle* (pp. 335–346). Éditions du GREF.

Uritescu, D., Mougeon, R., Nadasdi, T., & Rehner, K. (2004). Acquisition of the internal and external constraints of variable schwa deletion by French immersion students. In J.-M. Dewaele & R. Mougeon (Eds.), *Variation in the Interlanguage of Advanced Second Language Learners*. Special issue of *International Review of Applied Linguistics*, *42*(4), 349–364.

Valdman, A. (1981). Sociolinguistic aspects of foreigner talk. *International Journal of the Sociology of Language*, *28*, 41–52.

Valdman, A. (1982). Français standard et français populaire: Sociolectes ou fiction? *French Review*, *56*, 218–227.

Valdman, A. (1989). Classroom foreign language learning and language variation: The notion of pedagogical norms. In M. Eisenstein (Ed.), *The dynamic interlanguage: Empirical studies in second language variation* (pp. 261–277). Plenum.

Valdman, A. (1992). Authenticity, variation, and communication in the foreign language classroom. In C. Kramsch & S. McConnell-Ginet (Eds.), *Text and context* (pp. 79–97). D.C. Heath and Company.

Valdman, A., Pons, C., & Scullen, M. E. (2013). *Chez nous : Branché sur le monde francophone*. Pearson.

van Compernolle, R. (2009). Emphatic *ne* in informal spoken French and implications for foreign language pedagogy. *International Journal of Applied Linguistics*, *19*(1), 47–65.

van Compernolle, R. (2015). Native and non-native perceptions of appropriateness in the French second-person pronoun system. *Journal of French Language Studies*, *25*, 45–64.

van Compernolle, R. (2019). Constructing a second language sociolinguistic repertoire: A sociocultural usage-based perspective. *Applied Linguistics*, *40*(6), 871–893.

van Compernolle, R. A., Weber, A., & Gomez-Laich, M. P. (2016). Teaching L2 Spanish sociopragmatics through concepts: A classroom-based study. *Modern Language Journal*, *100*, 341–361.

van Compernolle, R., & Williams, L. (2009). Learner versus nonlearner patterns of stylistic variation in synchronous computer-mediated French: *Yes/No* questions and *nous* versus *on*. *Studies in Second Language Acquisition*, *31*, 471–500.

van Compernolle, R., & Williams, L. (2012a). Reconceptualizing sociolinguistic competence as mediated action: Identity, meaning-making, agency. *The Modern Language Journal*, *96*(2), 234–250.

van Compernolle, R., & Williams, L. (2012b). Teaching, learning, and developing L2 French sociolinguistic competence: A sociocultural perspective. *Applied Linguistics*, *33*, 184–205.

Vande Berg, M., Connor-Linton, J., & Paige, R. M. (2009). The Georgetown Consortium Project: Interventions for student learning abroad. *Frontiers: The Interdisciplinary Journal of Study Abroad*, *18*, 1–75.

VanPatten, B., & Benati, A. G. (2010). *Key terms in second language acquisition*. Continuum.

VanPatten, B., & Williams, J. (Eds.). (2014). *Theories in second language acquisition: An introduction* (2nd ed.). Routledge.

Varonis, E. M., & Gass, S. M. (1985). Non-native/non-native conversations: A model for negotiation of meaning. *Applied Linguistics*, *6*, 71–90.

Villeneuve, A.-J., & Auger, J. (2013). *'Chtileu qu'i m'freumereu m'bouque i n'est point coér au monne'*: Grammatical variation and diglossia in Picardie. *French Language Studies*, *23*, 109–133.

White, L. (2009). Grammatical theory: Interfaces and L2 knowledge. In W. Ritchie & T. Bhatia (Eds.), *The new handbook of second language acquisition* (pp. 49–68). Emerald.

Wilkinson, S. (1998). Study abroad from the participants' perspective: A challenge to common beliefs. *Foreign Language Annals*, *31*, 23–39.

Wilkinson, S. (2002). The omnipresent classroom during summer study abroad; American students in conversation with their French hosts. *Modern Language Journal*, *86*, 157–173.

Wolfram, W., & Christian, D. (1976). *Appalachian Speech.* Arlington, VA Center for Applied Linguistics.

Young, R. (1988). Variation and the interlanguage hypothesis. *Studies in Second Language Acquisition*, *10*, 281–302.

Young, R. (1991). *Variation in interlanguage morphology*. Peter Lang.

Young, R., & Bayley, R. (1996). VARBRUL analysis for second language acquisition research. In R. Bayley & D. Preston (Eds.), *Second language acquisition and linguistic variation* (pp. 253–306). John Benjamins.

Zahler, S. (2014). Variable subject doubling in spoken Parisian French. *University of Pennsylvania Working Papers in Linguistics*, *20*(1), 361–371.

Zahler, S. (2018). *The relationship between working memory and sociolinguistic variation in first and second languages: The case of Spanish subject pronouns* (Publication No. 10841553) [Doctoral dissertation, Indiana University Bloomington]. ProQuest Dissertations Publishing.

Zribi-Hertz, A. (1994). The syntax of nominative clitics in Standard and Advanced French. In G. Cinque, J. Koster, J.-Y. Pollock, L. Rizzi, & R. Zanuttini (Eds.), *Paths towards universal grammar: Studies in honor of Richard S. Kayne* (pp. 453–72). Washington, DC: Georgetown University Press.

Zribi-Hertz, A. (2011). Pour un modèle diglossique de description du français : Quelques implications théoriques, didactiques et méthodologiques. *Journal of French Language Studies*, *21*, 231–256.

Zuengler, J. (1991). Accommodating in native-nonnative interactions: Going beyond the "what" to the "why" in second-language research. In H. Giles, J. Coupland, & N. Coupland (Eds.), *Contexts of Accommodation* (pp. 223–244). Cambridge University Press.

Zwanenburg, W. (1965). *Recherches sur la prosodie de la phrase française*. Universitaire Pers.

Appendix A: Pilot study

This appendix provides full details on the pilot study introduced and briefly described in section 3.4. Data on several sociolinguistic variables were collected (*on/nous*, interrogative structures, null objects, and *ne*-retention) in this pilot study; since Dewaele's study (2004a) of different interlocutor types also focused on *ne*-retention, the results reported in this section will focus on this variable. For a review of the literature and discussion of empirical studies on *ne*-retention, see sections 4.1 through 4.3.

A.1 Participants

Recruitment of near-native speakers was carried out in three locations in France (Normandy, Alsace, and Paris). The locations were chosen due to availability of participants who were recruited via networking with friends and colleagues. Despite completing all tasks, I ultimately determined that the one speaker recruited in Paris did not meet the near-native criteria due to significant deviance from nativelike syntax and phonology (as was the case for two participants in Donaldson's study). This left six participants for this pilot study meeting the other minimum requirements of native speaker of English and at least three years of continuous residence in France (following Birdsong (1992) and Donaldson (2008)); see Table A-1.

Table A-1. Demographic information on near-native speakers (pilot study)

ID#	Age	Sex	CoB	AOI	AOE	AOA	LOR	Educ.	Profession	RR
1N	28	F	USA	11	20	23	5	MA	ESL teacher	Normandy
2N	43	M	USA	4	4	4, 27[122]	16[123]	BS	Engineer	Normandy
3N	42	F	Canada	6	27	36	6	BA	ESL professor	Alsace
4N	60	F	USA	12	20	24	36	PhD	ESL/Literature professor	Alsace
5N	86	M	UK	58	70	75	11	MA	retired	Normandy
6N	34	F	Philippines[124]	12	27	27	7	MA	student/ESL teacher	Normandy

Abbreviations: CoB = country of birth; AOI = age of first instruction in French; AOE = age of first major exposure to native speakers of French; AOA = age of continuous exposure to French (beginning of long-term stay in France); LOR = length of residency (years); Educ. = highest education level completed; RR = region of residence in France

Besides French, these speakers reported varying levels of fluency in German, Spanish, Italian, Mandarin, and Tagalog, though all speakers indicated that French was their most proficient L2. After participants expressed interest in the study following my initial e-mail contact, I then asked each participant to identify a near-native speaker (meeting the length of residence criteria) and a NS with whom they would be able to carry out a conversation in French. Two such dyads consisting of four Near-NSs (3N-4N; 5N-6N) were established in this way. Speakers 1N and 2N were not able to provide a Near-NS conversation partner, so a confederate chosen by me participated in these two Near-NS/Near-NS conversations. The confederate was an acquaintance of 2N and a friend of a colleague of 1N. The confederate's demographic information is listed below:

ID#	Age	Sex	CoB	AOI	AOE	AOA	LOR	Educ.	Profession
1A	27	F	USA	14	18	18	3[125]	MA	student/FLE professor

[122] Participant lived in France from the ages of 4-9 and again beginning at age 27.

[123] Only includes participant's current length of residency; does not include childhood residency.

[124] Participant was born in the Philippines but raised in the United States, reporting English as her L1 and Tagalog as an L2 learned in adulthood.

[125] In this case, LOR = total length of residence in France; at the time of the pilot study the confederate was living in the United States but had traveled to France to participate in a summer teaching program.

Concerning native speakers, for each dyad the Near-NSs were instructed to recruit a NS with whom they were comfortable speaking in French. As in Donaldson's (2008) study, the NS conversation partner was a friend, spouse, or other acquaintance with whom the Near-NS typically spoke French. In one case, the NS served as conversation partner for two Near-NSs (5N and 6N). All NSs (labeled as 1F through 5F) were born in France and had not spent more than a total of two years in a foreign country. Twelve dyads involving all of these speakers were therefore created: 1N-1A, 1N-1F, 2N-1A, 2N-2F, 3N-4N, 3N-3F, 4N-4F, 5N-6N, 5N-5F, 6N-5F.

A.2 Procedure

I met with each participant in their home (1N, 2N, 5N, 6N) or in their university office (3N, 4N). After explaining the basic procedure of the study, all participants signed consent forms and filled out the first part of a background questionnaire containing basic demographic information (derived from Donaldson, 2008; see Appendix E).

Afterward, the participants in the Near-NS/NS dyad seated themselves in a comfortable conversation setting. Three of the dyads took place in a sound lab on the university campus; while this setting may have been less informal than a residence or lounge area, it was a neutral setting for both participants and had the added advantage of eliminating any background noise from the recording. A small, digital audio recorder connected to a wired lapel microphone was placed near each speaker, who attached the microphone to his or her clothing in a comfortable position. The use of two recorders facilitated identification of each speaker in addition to making overlapping speech easier to understand during post-hoc analysis. A third recorder was also placed between the two speakers, serving as a backup in case of failure of one or both of the other recorders. Following Donaldson (2008), I instructed the participants to speak for at least 45 minutes and to simply "catch up" with each other, developing the conversation with any subsequent topic that came to mind;

that is, no conversational topics were prescribed or suggested. This also means that neither speaker was assigned a particular role in this production task, thus minimizing any perceived power imbalance between participants. When the participants indicated that they were ready to begin the conversation, I turned on all three recorders and left the room. After this conversation was finished, the participants took a short break, after which the Near-NS/Near-NS dyad completed the second conversation.

The remaining written tasks were then completed: the second part of the background questionnaire, focusing on language habits,[126] followed by the Acceptability Judgment Task. Most of the Near-NS participants were able to complete the AJT following the conversation tasks, though due to time constraints, two of them chose to complete the judgment task at a later date and mail their responses back to me (a procedure that Donaldson also allowed). I then debriefed each participant on the nature of the study and offered them 15 euros as a token of appreciation for their time.

A.3 Results: Acceptability Judgment Task and language security index

For the pilot study, due to time concerns, I did not have my NS interlocutors complete the same AJT, so *t*-tests on NS and Near-NS response patterns for the same groups are not possible. However, I can compare the average responses on each item according to its grammaticality, as determined by native speakers in Birdsong (1992), and compare these averages to that of near-natives in previous studies. Table A-2 provides these results (the higher the score, the more grammatical the item was rated).

[126] This part of the questionnaire was administered after the speaking task, since it contained questions asking the speaker to reflect on her use of the language and may have influenced the speaking task if it had been administered prior.

Table A-2. Acceptability judgment task results for Near-NSs in pilot study

Grammaticality of item	Pilot study	Birdsong (1992)	Donaldson (2008)
Grammatical items (*n* = 44)	3.68	3.72	3.64
Ungrammatical items (*n* = 26; denoted with *)	2.14	1.94	1.77
Questionable grammaticality (n = 7)	2.74	2.53	2.23

Based on these results, the six Near-NSs patterned similarly with the Near-NSs in Birdsong's and Donaldson's studies on their responses to grammatical items. As for ungrammatical items and those rated as questionable by native speakers, my Near-NS group had higher acceptance rates than in previous studies, though there is a clear distinction between clearly ungrammatical items and those that are not more uniformly rejected by native speakers.

For individual speakers, an additional analysis can be carried out by comparing Near-NS judgments with NS judgments in previous studies in order to determine the extent to which each Near-NS differed from NS averages. (This analysis is described in section 3.5.2.6; I reproduce discussion of the procedure here.) Since my NSs did not do the AJT, I took the average rating of the AJT results on each of the 76 AJT items for NSs reported in Birdsong (1992) and Donaldson (2008) and computed the difference for each Near-NS in the pilot study. I then averaged each difference to determine how much each speaker deviated from Birdsong's and Donaldson's NSs. I then subtracted the best theoretically possible score from each speaker's overall average (since items were judged as integers on a 1-5 scale and thus could not be identical to non-integer averages). In this way, a speaker who theoretically assigned the same rating as the average of these NSs for each item (rounded up or down depending on the non-integer average) would then have an average deviation of 0.0, while a speaker who consistently rated each item one point higher (or lower) than the NS average would have an average deviation of 1.0. The smaller the deviation, the

more closely the Near-NS matches with NS intuitions on complex grammar structures (see Table A-3).

Table A-3. Average deviation from Birdsong and Donaldson NSs on AJT (pilot study)

Speaker ID	Avg. deviation from NSs
1N	1.122
2N	0.851
3N	1.000
4N	0.880
5N	0.753
6N	0.842
Group average	0.908

The group average indicates a higher average deviation than the group average for Donaldson's near-native speakers (0.753). A two-tailed *t*-test revealed a significant difference in these scores (p = .03), suggesting that the Near-NSs in the pilot study are not as nativelike as Donaldson's near-natives. Note, however, that this comparison with NSs on grammaticality intuitions does not correlate in all cases with informally observed proficiency of the Near-NS speakers in the pilot study, based on overall fluency, phonological accuracy, and mastery of colloquial forms. Such discrepancies between internal grammar systems (competence) and external processes (performance) are not uncommon. Furthermore, response strategies may have skewed the results; Speaker 1N had a high bias for grammatical responses across the board, judging a large number of ungrammatical items as grammatical (with a *d*-prime bias approaching zero), despite the fact that her informally observed proficiency was among the highest of this participant group. Conversely, speaker 6N's oral production included numerous grammatical errors, yet she produced the second lowest average deviation on the AJT, lower than speaker 4N whose speech contained no observed grammatical errors. The AJT's focus on complex grammatical structures, while serving as a broad measure of grammatical competence, does not always correlate with more

nativelike spontaneous oral production. However, the AJT results for these Near-NSs suggest that, as a group, their proficiency may not be sufficiently advanced that they can be broadly compared to previous studies on near-native speakers, such as Birdsong (1992) and Donaldson (2008).

I also provide the results of the language security index (see section 3.5.1.4) for the Near-NSs in the pilot study (Table A-4). Recall that these are self-reported responses concerning the speaker's ability to read, write, speak, and hear French, as well as ability and motivation to pass as a native speaker (see questions #17-22 and #24 in Appendix E).

Table A-4. Measure of language security in Near-NSs (pilot study)

Speaker ID	**Abilities in French**					**Motivations in French**		**Security Index (1-5)**
	Read	Hear	Write	Speak	Accent rating	Pass as NS	Attempt to pass as NS	
1N	5	5	4	5	5	4	5	4.7
2N	4	4	3	5	4	3	3	3.7
3N	4	4	3	5	3	2	1	3.1
4N	5	4	4	5	4	3	4	4.1
5N	5	5	4	4	4	3	1	3.7
6N	5	4	4	4	3	1	1	3.1
Average	4.7	4.3	3.7	4.7	3.8	2.7	2.5	3.8

A.4 Results: *Ne*-retention

Upon completion of the study, I transcribed each of the conversations conducted by the six Near-NSs, and verbal negation contexts with possibility for variation were coded for the presence or absence of *ne* (eliminating non-variable expressions such as *n'est-ce pas* ('is it not') and ambiguous cases such as *on (n')est pas* ('one NEG is not')). Table A-5 summarizes the results for each Near-NS and NS, accompanied by overall percentages of retention.

Table A-5. Ne-retention in Near-NS group (pilot study)

Speaker status	Speaker ID	*ne* tokens	Total verbal negation contexts	% *ne*-retention
Near-NS	1N	0	146	0.0
	2N	21	120	17.5
	3N	9	108	8.3
	4N	29	65	45.6
	5N	35	91	38.5
	6N	66	95	69.5
	Overall Near-NS	**160**	**625**	**25.6**
	1A (interlocutor)	**5**	**96**	**5.2**
NS interlocutors	1F	0	54	0.0
	2F	18	61	29.5
	3F	1	25	4.0
	4F	7	40	17.5
	5F	8	85	9.4
	Overall NS	**34**	**265**	**12.8**
All speakers		**199**	**986**	**20.2**

It is immediately apparent that there is considerable *ne*-retention variation in Near-NSs (range = 69.5%), while NSs show much less variation overall (range = 29.5%). These ranges are similar to results obtained in previous L1 and L2 studies on *ne*-retention. Concerning NSs, overall *ne*-retention rates pattern within the ranges of recent NS surveys, slightly lower than Coveney (2002) at 19% while higher than Pooley (1996) at 7% and Hansen and Malderez (2004) at 8%. For Near-NSs, overall *ne*-retention is considerably lower than in most L2 studies, slightly higher than Donaldson's (2017) near-native speakers (22.4%), and similar to rates found mainly in learner subgroups that had previously had a long-term study abroad (cf. Sax's (2003) most advanced learners (29.7%); Howard's (2012) intermediate study-abroad group (26.4%)). There is also considerable overlap in the frequencies of Near-NSs and NSs; however, the between-group difference between Near-NS and NS *ne*-retention was highly significant ($\chi^2(1) = 18.3$, $p < .0001$).

Regarding individual speakers, Table A-6 compares *ne*-retention for each Near-NS with their AJT results and language security indices.

Table A-6. AJT scores, language security index, and ne-retention (pilot study)

Speaker ID	AJT avg. deviation	Language security index	% *ne*-retention
1N	1.122	4.7	0.0
2N	0.851	3.7	17.5
3N	1.000	3.1	8.3
4N	0.880	4.1	45.6
5N	0.753	3.7	38.5
6N	0.842	3.1	69.5

For AJT scores, the correlation with *ne*-retention appears to trend in the other direction than expected: the two speakers with the highest AJT deviation (1N and 3N) have the lowest *ne*-retention scores, while two of the three lowest AJT deviation scores (4N and 6N) have higher than average *ne*-retention. Furthermore, a higher language security index appears to correlate with lower *ne*-retention only in certain learners, as observed in speakers 1N (high security, low *ne*-retention) and 6N (low security, high *ne*-retention). Speaker 3N's low language security can be explained by her lack of motivation to pass as a native speaker; though her phonology is rather non-targetlike, possibly explaining this lack of motivation, her overall fluency is high. Conversely, speaker 4N's relatively high *ne*-retention, despite her high security and high grammatical accuracy, may be influenced by a combination of her professional status and age (as a highly accomplished late-career university professor); such factors may cause this highly proficient speaker to consciously choose to avoid informal variants, creating, as Dewaele (2007) cautions, a false presentation of incomplete sociolinguistic competence.

A.4.1 Ne-retention by interlocutor status

I now turn to the *ne*-retention results in the pilot study according to interlocutor type. As reported in section 3.4.1, I reproduce Table 3-2 as Table A-7 below, comparing this pilot study with the results from Dewaele (2004a).

Table A-7. Ne-retention rates in pilot study and comparison with Dewaele (2004a)

	Overall results (all speakers)			**L1 and L2 results divided by L1 status**			**L2 speakers: results across interlocutor type**		
	ne	Total neg	% *ne*	French status	*ne* / total neg	% *ne*	Interlocutor type	*ne* / total neg	% *ne*
Dewaele (2004a)	628	991	63.4	L1	N/A	36.3	L1 French	N/A	53.5
				L2	N/A	72.8	L2 French	N/A	75.5
Pilot: Near-NSs	199	986	20.2	L1	34/265	12.8	L1 French	93/320	29.1
				L2[127]	160/625	25.6	L2 French	67/305	22.0

It is immediately striking that the current study produced a far lower overall *ne*-retention rate (20.2%) than that of Dewaele (63.4%). These results can be explained in part by three factors. First, the pilot includes Near-NS participants all of whom appear to have a higher proficiency level than the non-native speakers identified in Dewaele and thus less likely to retain *ne*, assuming that *ne*-retention in informal contexts generally decreases as proficiency increases. Second, in Dewaele's study the topics of conversation were determined in advance, which included discussion of formal and informal topics modeled on Labov (1972), whereas the current study did not indicate specific topics. Finally, in Dewaele's study the researcher was one of the interviewers (and likely older than most of the participants), and the speakers did not necessarily know each other—both elements introducing higher levels of formality—whereas in this pilot study the researcher (me) was excluded from the conversation sphere before the start of the recorded conversation, and the

[127] These figures do not include the near-native confederate's totals (5 *ne* tokens in 96 negation contexts).

familiarity of the acquaintances with each other ranged from somewhat casual (acquaintances or friends of friends) to intimate (spouses), resulting in a comparatively lower level of formality.

While the aforementioned factors may also help to explain the lower occurrences of *ne* by NSs in this pilot study compared to NSs in Dewaele (2004a), the proficiency of the Near-NSs in the current study certainly accounts for a large portion of the gap between these speakers and the NNSs in Dewaele (indeed, the Near-NSs in this sample retain *ne* at a lower rate than Dewaele's NSs).

As noted in section 3.4.1, Dewaele found a statistically significant difference in L2 *ne*-retention in conversation with NSs (53.5%) compared to conversation with other learners (75.5%). The reverse trend obtained in the pilot study: Near-NSs retained *ne* at higher rates with NSs (29.1%) than with other Near-NSs (22.0%), a difference of 7.1% that was marginally significant ($\chi^2(1) = 4.13$, $p = .042$).

A.5 Pilot study variationist analysis

A variationist analysis of *ne*-retention in Near-NSs was also conducted for the pilot study. Note that this analysis was carried out in Goldvarb X with fewer linguistic factor groups, no sociostylistic factor groups, and no random intercepts compared to the variationist analyses of SA learners and Near-NSs in the large-scale study as reported in Chapter 4. While this analysis does not provide a direct comparison for Near-NS results in the large-scale study, some general similarities can be observed. Table A-8 lists the significant factor groups for *ne*-retention in Near-NSs (in order of significance) and Table A-9 provides the results of non-significant factor groups.

Table A-8. Factor groups selected as significant for ne-retention in Near-NSs (pilot study)

Factor Group	Factor	Weight	% *ne*-retention	N
Speaker	1N	---	0.0	0/146
	2N	.650	18.5	22/119
	3N	.848	8.1	9/111
	4N	.397	44.8	30/67
	5N	.353	39.6	36/91
	6N	.129	69.5	66/95
Lexicalization	lexicalized	.644	18.5	31/168
	not lexicalized	.422	41.9	132/315
Verb type	main	.424	37.5	109/291
	auxiliary	.430	48.1	25/52
	copula	.718	15.6	19/122
	modal	.356	55.6	10/18
Subject type	pronoun	.521	31.7	146/461
	full NP	.105	83.3	10/12
Relative pronoun?	*qui*	.168	61.5	8/13
	not *qui*	.511	33.0	155/470
TOTAL			33.7	163/483

Input: 0.728; Significance = 0.012

Table A-9. Non-significant factor groups in ne-retention in Near-NSs (pilot study)

Factor Group	Factor	% ne-retention	N
Following segment	vowel	29.6	68/230
	consonant	37.5	95/253
Post-verbal negator	*pas*	33.5	145/433
	others	36.0	18/50
Object pronoun?	present	53.3	8/15
	absent	33.1	155/468
Interlocutor L1	Near-NS	31.2	68/218
	NS	35.8	95/265
Gender	Male	27.6	58/210
	Female	38.4	105/273
Age group	Younger	29.8	97/325
	Older	41.7	66/158

As the significant factor group rankings indicate, individual speakers are the most significant source of variation. This is unsurprising, given the range of *ne*-retention rates in this small sampling of speakers, and it motivates the inclusion of a random intercept to account for

individual speaker variation. Otherwise, the most significant factors center around the verb (Lexicalization and Verb type). The remaining two significant factors concern the subject type, though the variants favoring *ne*-retention for these two factor groups (full NP and the relative pronoun *qui*) were produced in only 12 and 13 utterances respectively, or 5% of the total negation contexts. Concerning extralinguistic contexts, Interlocutor L1 was not selected as a significant factor influencing *ne*-retention for these speakers. Gender and Age are not significant either, though all these variables are conditioned by the small sample size (especially given that speaker 1N's data was not included for factor analysis due to categorical *ne*-retention).

As far as the selection of significant factors, these results diverge in certain ways from previous variationist analyses such as Regan (1996) and Sax (2003). Among all factors that were considered in each study, Subject type was significant in this pilot study, as in Regan and Sax. Lexicalization was also significant in Regan, while Verb type was significant in Sax. As for ordering of significant factors, Subject type was more significant for Regan's and Sax's learners, though Lexicalization does appear lower on Regan's list of significant factors. Interestingly (but perhaps unsurprisingly), Near-NSs in this pilot study pattern more like the native speakers rather than the lower-level learners in these other studies; Lexicalization was the most significant linguistic factor in Sax's NS interviewers and NS control group, as well as in Ashby's (1981) native speaker study.

Linguistic factors thus seem to have marginalized the interlocutor effect that was found in the cross-sample significant difference in *ne*-retention, and in the variationist analysis, the difference across interlocutor type is lowered compared with the overall results which contain the data of the speaker (1N) with categorical deletion. It is possible that the status of the interlocutor

did have an effect on this sample of Near-NSs, but that the effect is too small to have any significant impact in their choice of deletion or retention as measured in a variationist analysis.

A.6 Limitations in the pilot study

Certain limitations can be raised about the participant population, methodology, and results of this pilot study, all taken into account when designing the full-scale study. First, the demographic characteristics of some of the pilot study participants warrant discussion. Speaker 2N had spent several years of his childhood in France; though he offered in the debriefing that he had forgotten "all" of his French when he moved back to the United States at age 9, the potential effect of this childhood stay in France may influence his speech in ways that cannot allow for comparison with other learners who began acquiring French after the putative "critical period" in language acquisition. At the other end of the spectrum, speaker 5N participated in this study at age 86 after having had his first significant exposure to native French speakers at the age of 70; though this speaker's age does not *a priori* exclude him as a potential near-native speaker of French (indeed, one of Donaldson's near-native speakers was 70 years old), his age and (perceived) social divergence from his near-native interlocutor (6N; a female student in her thirties) may have influenced the latter's use of informal features. Speaker 6N also produced a sufficient number of grammatical errors in her speech that it may be inaccurate to categorize her as a near-native speaker, regardless of her performance on the AJT. Furthermore, 6N's substantially high *ne*-retention rate may be partly explained by her responses in the language habits questionnaire, indicating a learner who rarely passes as, and rarely attempts to pass as, a native speaker. Finally, the significant difference between this Near-NS group and Donaldson's near-native group regarding the AJT average deviation scores suggests that more careful selection of participants is in order.

The results from the pilot study were also potentially affected by other logistical factors. Due to time constraints, it was not always possible to control which speaker would begin with which interlocutor, thus introducing possible ordering effects. Furthermore, the location of the recordings was also dependent on the availability of the participants, who could not all arrange to meet in their homes. Finally, the results obtained from this study must be interpreted cautiously due to the small sample of participants and the marginally significant *p* values for *ne*-retention differences across interlocutor types. Ideally, at least as many participants ($n = 10$) as those in Donaldson (2008) would provide for more direct comparisons with his tasks and results, and a larger difference in the *ne*-retention rates, as measured by statistical significance, would similarly allow for a more robust observation of potential interlocutor effects.

Appendix B: Pragmatic particles in informal discourse

This appendix continues discussion initiated in section 3.6.2 concerning the specific pragmatic particles produced by the participants in the current study, as part of a method for determining the formality of these participants' conversations. Recall that, in addition to *enfin* and *hein* observed in Donaldson's (2008) informal features, I include *quoi, genre, machin,* and *bref* in the distribution of pragmatic particles for these participants. In the following sections, I briefly introduce these particles before providing the distribution of all pragmatic particles for each participant. For more details on *enfin*, see Beeching (2001, 2011), Bertrand and Chanet (2005), and Donaire (2013); for details on *hein*, see Beeching (2004) and Valdman (1982). For a recent treatment of pragmatic particles in L2 French discourse, see Reaves (2020).

B.1 Pragmatic particles *quoi*, *genre*, *machin*, *bref*

Quoi functions as an interrogative pronoun, among other uses, but in informal speech it can appear at the end of an utterance to serve a variety of communicative functions: to emphasize what the speaker has just said, to indicate that the speaker is evaluating her own utterance in some way (including expressing hesitation or tentativeness about what she has just said), or to invite the interlocutor to make an assumption of the speaker's position on a particular topic (Beeching, 2004; Chanet, 2001).

(1) *c'est juste, il faut avoir les compétences quoi* (SaE)
'it's just, you have to be qualified, you know'

Quoi can be somewhat analogous to discourse particles in English such as "like" or "you know."

Genre (literally, "gender" or "kind") often appears as in *du genre* ('of the sort'), though it can appear in invariable form without a determiner. In informal French, *genre* can have a similar function as discourse "like" in English, such as introducing an example, an approximation, or

direct discourse (Fleischman & Yaguello, 2004). Its function as a discourse marker seems to have been adopted relatively recently, and it is used most frequently by young speakers (Isambert, 2016; Hennecke, 2017). In one context from my corpus, a speaker imagines a hypothetical situation of being introduced to her seven-year-old stepdaughter this year, rather than having actually met her stepdaughter four years ago when the latter was much younger:

(2) *si je l'avais rencontrée genre cette année, ça aurait été moche* (2L)
'if I had met her, like, this year, that would have been rough'

In this context, *genre* is used to introduce an example (*cette année*).

Machin (literally, "thing" or, more informally, "thingamajig") often appears as a "vague" noun similar to *truc* ('thing'), bare or with a determiner. It also has a discourse function as a "general extender" (Overstreet, 2005; Mihatsch, 2006) that is invariable (no determiner or plural inflection) and used after the speaker has cited a list of multiple items, in the same vein as *et cetera*:

(3) *…je veux être toute blanche, très très traditionnelle, et à l'église, machin machin* (CaF)
(talking about her ideal wedding): 'I want to be all in white, very very traditional, and at the church, et cetera, et cetera'

More rarely, *machin* can also appear elsewhere as a particle similar to *genre* (Béguelin & Corminboeuf, 2017). In my data, *machin* was uttered by two of my speakers (CaF and 2P) as a vague noun and as a general extender; the former tokens were considered as informal vocabulary, while the latter tokens were considered as pragmatic particles.

Bref (literally, "brief") is often used in conjunction with *enfin* to indicate the end of an utterance in which the speaker believes she has spoken for too long or does not wish to belabor her point (analogous to English "anyway"), and may be followed by a short summary statement.

(4) *mes amis, ils aiment pas forcément, ils vont dire « vas-y, casse-toi », enfin bref…* (FrE)
'my friends, they don't necessarily like [that I root for England], they'll say "come on, get the hell out of here," anyway…'

B.2 Distribution of pragmatic particles in speaker groups

Recall that for Tables 3-21 through 3-25, and for Table 3-29, if the speaker used any of the aforementioned pragmatic particles, the category of "Pragmatic particles" was marked with a plus sign, indicating that this speaker had demonstrated use of this informal feature. In this section, Tables B-1 and B-2 show the distribution of these pragmatic particles, and the number of tokens for each particle, uttered by Near-NSs in Pau and Lille, respectively. When no particles were produced, no number is indicated. For the Interlocutor column, *E* indicates the conversation with the near-native speaker (Lille) or the English-identity bilingual (Pau); *F* indicates the conversation with the native speaker (Lille) or the French-identity bilingual (Pau).

Table B-1. Pragmatic particles in Near-NSs (Pau)

Particle	Inter-locutor	1P	2P	3P	4P	5P	6P	7P	8P	9P	10P
enfin	E		8	3		2					
	F	2	11	10		5					
hein	E	2	3	1				1		2	
	F		1								
quoi	E	1	7					5		1	
	F		10	5			2	5	1		
genre	E										
	F										
machin	E		1								
	F										
bref	E		4								
	F		3								

Table B-2. Pragmatic particles in Near-NSs (Lille)

Particle	Inter-locutor	1L	2L	3L	4L	5L	6L	7L	8L	9L
enfin	E		22			1			6	
	F		23			6			12	
hein	E				1	1				1
	F					1		2		
quoi	E							38	4	5
	F				2	3		22	5	6
genre	E		2							
	F									
machin	E									
	F									
bref	E									
	F									

Table B-3 provides the pragmatic particles for all interlocutors, including the interlocutors for the Near-NSs at both sites and the interlocutors for the SA learners (AmE and SoF). For the bilinguals (Ch, Fr, Th), tokens are broken down according to whether they were uttered under their English and French identity; for the remaining interlocutors, recall that SaE and JeE are near-native (L1 English) speakers, while CaF and KeF are native French speakers.

Table B-3. Pragmatic particles for all interlocutors

Particle	AmE	SoF	ChE	ChF	FrE	FrF	ThE	ThF	SaE	JeE	CaF	KeF
enfin	19	13	30	41	70	86	45	7			197	13
hein			1	1					1		10	6
quoi	20	4	14	22	16	14			94		37	55
genre					6	15			1	4		
machin											11	
bref					2						2	

As these tables show, *enfin* is generally the most common of these particles, appearing in the speech of all native speakers, while *quoi* is also frequent in both near-native and native speech. Near-native speakers also seem to adopt and favor certain particles over others, to a greater degree

than NSs. Most striking are 2L's use of *enfin*, 7L's use of *quoi*, and SaE's use of *quoi*. Finally, note that none of the SA learners produced any tokens of the pragmatic particles described in this section—an unsurprising outcome for this proficiency level (see Reaves (2020: 236) for similar learner results with the particles *enfin*, *ben*, and *voilà*).

Appendix C: Written topic prompts for SA learners

(Interlocutor descriptions and English translations seen here are provided for information purposes only and were not seen by the participants)

With near-native interlocutor (AmE):
Journée typique dans le programme d'été
Les familles d'accueil
Voyages en France/en Europe
Parler français à Pau/en France
Vivre en France : ce que j'aime/ce qui est difficile

With native interlocutor (SoF):
Etudes universitaires / jobs / projets pour l'avenir
Ma famille
Endroits à visiter dans ma région
Les films français / la musique française
Festivals d'été autour de Pau

With other SA learner:
La nourriture française : ce que j'aime/je déteste manger
La nourriture qui me manque aux Etats-Unis
Les transports à Pau (bus, vélo, marcher, etc.)
Endroits préférés à Pau (restaurants, cafés, parcs, etc.)
Mes buts/objectifs pour la deuxième session d'été

English Translations:
With near-native interlocutor (AmE):
Typical day in the summer program
Host family experiences
Travels in France/Europe
Speaking French in Pau/in France
Pleasures/challenges of living in France

With native interlocutor (SoF):
University studies / job experiences / future plans
Family at home
Places to visit in your region
French movies/music
Summer festivals in or near Pau

With other SA learner:
Favorite/least favorite foods in France
Foods you miss from home
Transportation in Pau (bus, bicycle, walking, etc.)
Favorite places in Pau (restaurants, cafés, parks, etc.)
Goals for the second summer session

Appendix D: Proficiency measure (*c*-test)

Participants were given a sheet of paper containing the first two paragraphs only; the expected answers and translations in English are provided here for information purposes and were not seen by the participants.

1. Un livre qui prétend introduire des aspects de la culture française ne serait pas complet sans un chapitre sur les beaux-arts. En fa__, de nomb_____ touristes vo__ en Fra____ dans l'inte_____ d'admirer s__ chefs-d'œu____ de pein_____, d'archit_____ et d__ sculpture. Q___ n'a p___ entendu par____ du Louvre ? d__ la cath_____ Notre-Dame d__ Paris ? des scul_____ de Rodin ? No___ ne pou_____ pas vo___ présenter u__ étude e__ profondeur d___ beaux-ar___ en Fra____.

2. Quand on revient d'un voyage dans un pays étranger, la première chose dont on se souvient est presque toujours la cuisine: non seul_____ la nourr_____ mais au____ la fa____ de l__ préparer, d__ la man____, les heu____ des re____, tous l___ rites q___ les accomp_____ et q___ caractérisent l__ gens d__ pays mi____ que n'imp_____ quel au____ aspect d__ la v____. En Fra____, la gastr_____ est particul_____ importante, c___ c'est u__ véritable art ; et il ne s'agit pas d'un art pratiqué par un petit nombre de spécialistes, mais d'un art auquel participe toute la population.

Expected Answers
1. Un livre qui prétend introduire des aspects de la culture française ne serait pas complet sans un chapitre sur les beaux-arts. En fait, de nombreux touristes vont en France dans l'intention d'admirer ses chefs-d'œuvre de peinture, d'architecture et de sculpture. Qui n'a pas entendu parler du Louvre ? de la cathédrale Notre-Dame de Paris ? des sculptures de Rodin ? Nous ne pouvons pas vous présenter une étude en profondeur des beaux-arts en France.

2. Quand on revient d'un voyage dans un pays étranger, la première chose dont on se souvient est presque toujours la cuisine: non seulement la nourriture mais aussi la façon de la préparer, de la manger, les heures des repas, tous les rites qui les accompagnent et qui caractérisent les gens du pays mieux que n'importe quel autre aspect de la vie. En France, la gastronomie est particulièrement importante, car c'est un véritable art; et il ne s'agit pas d'un art pratiqué par un petit nombre de spécialistes, mais d'un art auquel participe toute la population.

Translation in English:
1. A book that purports to introduce aspects of the French culture wouldn't be complete without a chapter on Fine Arts. Actually, many tourists go to France in order to admire its painting, architecture and sculpture masterpieces. Who has never heard about the Louvre? About the cathedral Notre-Dame de Paris? About Rodin's sculptures? We cannot introduce you to an in-depth study of the Fine Arts in France.

2. When we come back from a trip to a foreign country, the first thing we remember is almost always the cuisine: not only the food, but also the way of cooking it, of eating it, the meal hours, all the customs that accompany them and that characterize the people of the country more than any other aspect of life. In France, gastronomy is particularly important, because it is a true art; it is not about an art practiced by a small number of specialists, but an art in which the whole population participates.

Appendix E: Language background questionnaire

Please take a moment to fully answer the following questions regarding your language experience. All information is confidential; you may choose not to answer any of the questions, but we encourage you to be as accurate as possible in your answers.

1. Gender: Male / Female (circle one)

2. Age:

3. Highest academic degree earned (or current year of undergraduate study):

4. Profession:

5. Country of birth:

6. Native language(s):

7. Other languages you have studied or received significant exposure to:
   Language Length of study/exposure

8. How old were you when you began studying French?

9. How did you first start learning French? (i.e., school, private tutor, friends/relatives)

10. If you began learning French in school, were your French teachers native speakers, or did they learn French as a foreign language?

11. How old were you when you had your first extensive contact with native speakers of French?

12. How long have you studied or lived in a French-speaking country (including your current situation and past experiences)?

	Country/region	Purpose of stay	Approximate dates of stay
Experience 1			
Experience 2			
Experience 3			
Experience 4			

13. Why did you move to France? or: What is the reason for your current stay in France?

14. Do you have, or have you had, a spouse or partner who is a native speaker of French? If so:
    How long has the relationship lasted?
    What language do/did you speak together?

15. Do other members of your family (parents, siblings, children, etc.) speak French?

16. What are your principal motivations for learning French?

Language habits

17. Take a moment to consider your ability to READ in French, and circle the most appropriate response:

    (1) I can recognize a few words when I see them but can't really read the language.
    (2) I am sometimes able to understand the general meaning of a sentence, although there are many words I don't know.
    (3) I recognize and understand about half of what I read.
    (4) I recognize and understand most of what I read, although there are still sometimes words I don't know.
    (5) I recognize and understand everything or nearly everything I read, and I rarely see words I don't know.

18. Take a moment to consider your ability to UNDERSTAND WHAT YOU HEAR in French, and circle the most appropriate response:

    (1) I can recognize a few words when I hear them but can't really understand what is being said.
    (2) I am sometimes able to understand the general meaning of a sentence, although there are many words I don't know.
    (3) I recognize and understand about half of what I hear.
    (4) I recognize and understand most of what I hear, although there are still sometimes words I don't know.
    (5) I recognize and understand everything or nearly everything I hear, and I rarely hear words I don't know.

19. Take a moment to consider your ability to WRITE in French, and circle the most appropriate response:

    (1) I can't really write in the language, although I know a few words.
    (2) I can write very basic sentences in the language.
    (3) I can write a paragraph in the language, although there may be errors.
    (4) I can write well in the language, although there may occasionally be errors.
    (5) I can write the language as well as or nearly as well as a native speaker of the language.

20. Take a moment to consider your ability to TALK in French, and circle the most appropriate response:

    (1) I can't really speak the language, although I know a few words.
    (2) I can say or ask for very basic things, and generally make myself understood.
    (3) I can say or ask for a many things, and usually make myself understood.
    (4) I can say or ask for most things, and do not usually have trouble communicating.
    (5) I can say or ask for anything as effectively or nearly as effectively as a native speaker.

21. Take a moment to think about your ACCENT in French, if you have one, and circle the most appropriate response:

    (1) I am unable to pronounce most of the words in the language.
    (2) I have a strong foreign accent, and people often do not understand what I say in the language.
    (3) I have a noticeable foreign accent, but people generally understand what I say in the language.
    (4) I have a slight foreign accent, but people usually understand me easily.
    (5) I have no foreign accent or almost no foreign accent, and most people would think I am a native speaker.

22. Have you ever passed as a native speaker of French? Circle the most appropriate response:

    (*1) Never*
    (2) Rarely
    (3) Sometimes
    (4) Most of the time
    (5) Always or almost always

23. If you sometimes pass as a native speaker of French, in what situations (or with whom) has this occurred?

24. Have you ever tried consciously to pass as a native speaker of French? Circle the most appropriate response:

    (*1) Never*
    (2) Rarely
    (3) Sometimes
    (4) Most of the time
    (5) Always or almost always

25. If you fail to communicate when speaking in French to a native speaker (in this case, spouse/partner, friend, colleague, or acquaintance), do you ever switch to speaking English?

    (*1) Never*
    (2) Rarely
    (3) Sometimes
    (4) Most of the time
    (5) Always or almost always

26. If you fail to communicate when speaking in French to a native speaker (in this case, other speakers in the community, such as bank tellers, shopkeepers, or other professionals, or passersby on the street), do you ever switch to speaking English?
    (*1) Never*
    (2) Rarely
    (3) Sometimes
    (4) Most of the time
    (5) Always or almost always

27. If you are speaking to native English speakers or native speakers of a language other than French, do you feel that it is necessary to speak French when native French speakers are part of the conversation?

    (*1) Never*
    (2) Rarely
    (3) Sometimes
    (4) Most of the time
    (5) Always or almost always

<u>Cultural integration</u>

28. For each of the items below, please circle the response that corresponds to the amount of time you spend doing the activity IN FRENCH.

    —watching television in French:
    1. never
    2. rarely (once a month or less)
    3. sometimes (once or twice a week)
    4. daily

    —reading newspapers or the news online in French:
    1. never
    2. rarely (once a month or less)
    3. sometimes (once or twice a week)
    4. daily

—reading novels in French:
1. never
2. rarely (once a month or less)
3. sometimes (once or twice a week)
4. daily

—listening to the radio or playing music in French:
1. never
2. rarely (once a month or less)
3. sometimes (once or twice a week)
4. daily

—watching movies in French:
1. never
2. rarely (once a month or less)
3. sometimes (once or twice a week)
4. daily

—socializing with native speakers of French:
1. never
2. rarely (once a month or less)
3. sometimes (once or twice a week)
4. daily

29. Which language(s) do you use in the following situations? Please respond to all the situations that apply to you at the moment:

—at home (with host family, if applicable):
1. English only
2. Mostly English with occasional French
3. About half English and half French
4. Mostly French with occasional English
5. French only
6. other: _____

—with friends (native French speakers) living in France:
1. English only
2. Mostly English with occasional French
3. About half English and half French
4. Mostly French with occasional English
5. French only
6. other: _____

—with friends (native English speakers or native language other than French) living in France:

1. English only
2. Mostly English with occasional French
3. About half English and half French
4. Mostly French with occasional English
5. French only
6. other: _____

—at clubs/groups/church/other community functions in France:

1. English only
2. Mostly English with occasional French
3. About half English and half French
4. Mostly French with occasional English
5. French only
6. other: _____

—with family and friends living elsewhere:

1. English only
2. Mostly English with occasional French
3. About half English and half French
4. Mostly French with occasional English
5. French only
6. other:

Appendix F: Acceptability Judgment Task for near-native speakers

(*Beginning on the next page, participants saw each item in one of three randomized versions. The English translations and the grammaticality indicators (asterisk or question mark) are provided for information purposes only and were not seen by the participants).*

Jugements de phrases

Dans ce questionnaire, il s'agit de vos intuitions à propos d'un certain nombre de phrases en français. Lisez les consignes suivantes avant de commencer.

1. Nous nous intéressons à votre première réaction – c'est-à-dire à votre intuition sans réflexion – suscitée par chacune des phrases. Votre jugement portera sur le degré d'acceptabilité de chaque phrase.

2. Il n'y a pas de bonnes ni de mauvaises réponses; l'essentiel, c'est votre jugement instantané et immédiat sur chaque phrase.

3. Ne prenez donc absolument pas le temps de réfléchir à des règles de grammaire, etc. Ne demandez pas l'avis d'une autre personne.

4. Par contre, prenez le temps de lire chaque phrase une fois à haute voix pour que vous l'entendiez à l'oral.

5. Vous répondrez en mettant une croix à côté de la réponse qui convient le mieux à votre jugement sur la phrase. Vous choisirez parmi les réponses suivantes:

   ___A. pas du tout acceptable; je ne la dirais pas
   ___B. acceptable dans de rares contextes
   ___C. acceptable dans à peu près la moitié des contextes
   ___D. acceptable dans la plupart des contextes
   ___E. tout à fait acceptable; je la dirais

6. Une fois la réponse donnée, avancez à la question suivante. Ne revenez jamais en arrière pour changer une réponse préalable; ne changez jamais la réponse une fois qu'elle est marquée.

7. Nous vous prions de répondre aux questions en respectant l'ordre dans lequel elles sont présentées.

8. Nous vous signalons que les pages sont imprimées recto-verso.

Items

1. *Albert finira son travail bientôt.*
   "Albert will finish his work soon."

2. **Lucie a donné à Henri des fleurs.*
   "Lucie gave to Henri some flowers."

3. **Thomas a pris avant la fête une douche.*
   "Thomas took before the party a shower."

4. *Diane a placé des fleurs dans sa chambre.*
   "Diane put some flowers in her room."

5. *Antoinette a traversé rapidement la rue.*
   "Antoinette crossed rapidly the street."

6. **Diane a placé dans sa chambre des fleurs.*
   "Diane put in her room some flowers."

7. *Lucie a donné des fleurs à Henri.*
   "Lucie gave some flowers to Henri."

8. *Les garçons regardent la télévision avec intérêt.*
   "The boys watch the television with interest."

9. *Jeanne mange souvent de la crème glacée.*
   "Jeanne eats often ice cream."

10. *Marie a descendu prudemment les marches.*
    "Marie descended prudently the stairs."

11. *Antoinette a traversé la rue rapidement.*
    "Antoinette crossed the road rapidly."

12. *Albert finira bientôt son travail.*
    "Albert will finish soon his work."

13. *Jeanne mange de la crème glacée souvent.*
    "Jeanne eats ice cream often."

14. *Les garçons regardent avec intérêt la télévision.*
    "The boys watch with interest the television."

15. *Marie a descendu les marches prudemment.*
    "Marie descended the stairs prudently."

16. *Thomas a pris une douche avant la fête.*
"Thomas took a shower before the party."

17. Maurice a vu un certain lit. —> **Maurice en a vu un certain.*
"Maurice saw a certain bed. —> Maurice—of beds—saw a certain one."

18. Charles a vu un grand lit. —> *Charles en a vu un grand.*
"Charles saw a big bed. —> Charles—of beds—saw a big one."

19. Elle a lu ce livre. —> **Elle en a téléphoné à l'auteur.*
"She read this book. —> She—of it—phones the author."

20. La préface de ce livre est flatteuse. —> *La préface en est flatteuse.*
"The preface of this book is flattering. —> The preface of it is flattering."

21. Elle a lu ce livre. —> *Elle en aime l'auteur.*
"She read this book. —> She—of it—likes the author."

22. Le président de cette compagnie se paie un bon salaire. —> ?**Le président s'en paie un bon salaire.*
"The president of this company pays himself a good salary. —> The president of it pays himself a good salary."

23. La préface de ce livre m'agace. —> **La préface m'en agace.*
"The preface of this book irritates me. —> The preface of it irritates me."

24. Le dénouement de cette tragédie se devine déjà. —> *Le dénouement s'en devine déjà.*
"The ending of this tragedy can be guessed already. —> The ending of it can be guessed already."

25. La doublure de ce veston se lave en 10 minutes. —> *La doublure s'en lave en 10 minutes.*
"The lining of this jacket can be cleaned in 10 minutes. —> The lining of it can be cleaned in 10 minutes."

26. ?*C'était un étonné candidat qui a perdu aux élections.*
"It was a surprised candidate who lost the election."

27. **Le connu romancier vient d'arriver.*
"The known novelist just arrived."

28. *Le très-connu Marcel Proust vient d'arriver.*
"The well-known Marcel Proust just arrived."

29. *C'était un étonné Giscard qui a perdu aux élections.*
"It was a surprised Giscard who lost the election."

30. **C'était un étonné candidat.*
"It was a surprised candidate."

31. ?*Le très-connu romancier vient d'arriver.*
"The well-known novelist just arrived."

32. **Qui crois-tu qui rendra visite à Marc?*
"Who do think that [nominative case] will visit Marc?"

33. **Qui crois-tu que rendra visite à Marc?*
"Who do you think that [accusative case] will visit Marc?"

34. *Que dis-tu que qui a acheté?*
"What do you say that who bought?"

35. **Qui dis-tu qui a acheté quoi?*
"Who do you say that bought what?"

36. **Qui disais-tu qu'a épousé Laure?*
"Who did you say that [accusative case] married Laure? OR: Who did you say that Laure married?"

37. **Qui disais-tu qui a épousé Laure?*
"Who did you say that married Laure?"

38. **Que dis-tu qu'a acheté qui?*
"What do you say that [accusative case] bought whom? OR: What do you say that who bought?"

39. *Qui disais-tu a épousé Laure?*
"Who did you say married Laure?"

40. *Qui crois-tu rendra visite à Marc?*
"Who do you think will visit Marc?"

41. *Que dis-tu que Marie a acheté?*
"What do you say that Marie bought?"

42. ?*Elle a oublié deux matinées de libre(s).*
"She forgot two mornings free."

43. **Elle a les lettres d'écrites.*
"She has the letters written."

44. *Elle a une matinée de libre.*
"She has a morning free."

45. **J'ai trouvé un problème de résolu.*
"I found a problem resolved."

46. *Elle a les matinées de libre(s).*
"She has the mornings free."

47. *Elle a deux lettres d'écrites.*
"She has two letters written."

48. *Elle a la matinée de libre.*
"She has the morning free."

49. ?*Deux matinées de libre(s) étaient marquées sur son calendrier.*
"Two mornings free were marked on her calendar."

50. **Elle a oublié une lettre d'écrite.*
"She forgot a letter written."

51. **Deux lettres d'écrites étaient retrouvées dans sa chambre.*
"Two letters written were found in his room."

52. *Encore un problème de résolu.*
"Another problem resolved."

53. *Elle a une lettre d'écrite.*
"She has a letter written."

54. **Elle a oublié deux lettres d'écrites.*
"She forgot two letters written."

55. *Elle a deux matinées de libre(s).*
"She has two mornings free."

56. *Elle a oublié une matinée de libre.*
"She forgot a morning free."

57. **Elle a la lettre d'écrite.*
"She has the letter written."

58. Qui est Victor Hugo? *C'est un grand écrivain du XIXème siècle.*
"Who is Victor Hugo? He's a great writer of the 19th century."

59. *Bernard Pivot, c'est un intellectuel qui n'a peur de rien.*
"Bernard Pivot, he is an intellectual who is afraid of nothing."

60. *Marie voulait dire à Jean qu'il était un génie que tout le monde respecte.* [*il* se réfère à Jean]
"Marie wanted to tell Jean that he was a genius that everyone respects. [*he* refers to Jean]"

61. *Sophie nous a dit qu'elle était malade hier soir.*
"Sophie told us that she was sick last night."

62. *Marie a dit de Jean que c'est un génie.*
"Marie said of Jean that he is a genius."

63. *Marie voulait dire à Jean que c'était un génie que tout le monde respecte*. [*ce* se réfère à Jean]
"Marie wanted to tell Jean that he was a genius that everyone respects. [*he* refers to Jean]"

64. **Bernard Pivot, il est un intellectuel qui n'a peur de rien.*
"Bernard Pivot, he is an intellectual who is afraid of nothing."

65. **Sophie nous a dit que c'était malade hier soir.*
"Sophie told us that she/it was sick last night."

66. *Marie a dit de Jean qu'il est un génie.*
"Marie said of Jean that he is a genius."

67. Qui est Victor Hugo? **Il est un grand écrivain du XIXème siècle.*
"Who is Victor Hugo? He is a great writer of the 19th century."

68. ?*Cette maison s'est achetée d'elle-même.*
"This house bought itself."

69. **Cette maison a été vendue d'elle-même.*
"This house was sold by itself."

70. **Cette pièce se balaie d'elle-même.*
"This room sweeps itself."

71. *?*Cette voiture se lave d'elle-même.*
"This car washes itself."

72. **Ce tire-bouchon s'utilise de lui-même.*
"This corkscrew works by itself."

73. *Ces formules se mémorisent d'elles-mêmes.*
"These formulas memorize themselves."

74. *Cette maison s'est vendue d'elle-même.*
"This house sold itself."

75. *Les langues romanes s'acquièrent d'elles-mêmes.*
"Romance languages learn themselves."

76. *Ces théorèmes s'apprennent d'eux-mêmes.*
"These theorems learn themselves."

Appendix G: Non-significant factor groups in variationist analyses

The following tables provide details on the non-significant factor groups for the variationist analyses conducted on *ne*-retention (Chapter 4) and subject doubling (Chapter 5). For each table in this appendix, the corresponding table for that speaker group's significant factor groups is also indicated in the table's title.

G.1. Non-significant factor groups for *ne*-retention

Table G-1. Non-significant factor groups in study-abroad learners (ne-retention); cf. Table 4-21

Factor Group	Factor	% ne-retention	N
Object clitic	clitic	78.9	15/19
	no clitic	55.9	208/372
Relative clause	relative clause	55.6	5/9
	no relative clause	57.1	218/382
Reinforcing adverb	adverb	66.7	2/3
	no adverb	57.0	221/388
Subject type	pronoun	55.9	203/363
	full NP	73.9	17/23
	qui	60.0	3/5
Negator	*pas*	57.1	210/368
	other	56.5	13/23
Topic	serious	60.0	24/40
	not serious	56.7	199/351
Tu/vous	*tu*	62.2	156/251
	vous	40.2	37/92
	unknown	62.5	30/48
Interlocutor L1	L1 French	49.3	70/142
	L2 French	61.4	153/249
Gender	male (n = 4)	57.9	154/266
	female (n = 2)	55.2	69/125
Conversation portion	0-5 min	50.0	39/78
	5-10 min	61.6	61/99
	10+ min	57.5	123/214
TOTAL		57.0%	223/391

Table G-2. Non-significant factor groups for SA interlocutors (ne-retention); cf. Table 4-22

Factor Group	**Factor**	**% ne-retention**	**N**
Hiatus	hiatus	28.6	2/7
	no hiatus	6.6	20/303
Object clitic	clitic	5.3	1/19
	no clitic	7.2	21/291
Verb type	auxiliary	3.7	1/27
	main	9.1	15/164
	copula	5.0	5/101
	modal	5.6	1/18
Clause type	subordinate	12.0	3/25
	main	6.7	19/285
Reinforcing adverb	adverb	0.0	0/5
	no adverb	7.2	22/305
Negator	other	17.1	7/41
	pas	5.6	15/269
Topic	serious	10.8	4/37
	not serious	6.6	18/273
Quoted speech	not formal	0.0	0/4
	not quoted	7.2	22/306
Conversation portion	0-5 min	8.9	5/56
	5-10 min	6.4	3/47
	10+ min	6.8	14/207
TOTAL		7.1%	22/310

Table G-3. Non-significant factor groups for Near-NSs in Lille (ne-retention); cf. Table 4-23

Factor Group	Factor	% ne-retention	N
Object clitic	clitic	12.0	3/25
	no clitic	15.5	126/812
Clause type	main clause	14.1	104/735
	subordinate clause	24.5	25/102
Lexicalization	not lexicalized	17.7	101/572
	lexicalized	10.6	28/265
Negator	*pas*	15.3	105/688
	other	16.1	24/149
Quoted speech	formal	20.0	1/5
	not quoted	15.5	125/804
	not formal	7.9	3/28
Interlocutor L1	L1 French	15.8	79/501
	L2 French	14.9	50/336
Age	younger	14.0	74/527
	older	17.7	55/310
Gender	male (n = 4)	19.9	87/437
	female (n = 5)	10.5	42/400
Conversation portion	0-5 min	14.2	16/113
	5-10 min	8.9	8/90
	10+ min	16.6	105/634
TOTAL		15.4%	129/837

Table G-4. Non-significant factor groups for Near-NSs in Pau (ne-retention); cf. Table 4-24

Factor Group	Factor	% ne-retention	N
Object clitic	clitic	19.2	14/73
	no clitic	20.7	238/1151
Clause type	main clause	18.3	188/1029
	subordinate clause	32.8	64/195
Reinforcing adverb	adverb	25.4	16/63
	no adverb	20.3	236/1161
Negator	*pas*	19.4	202/1041
	other	27.3	50/183
Tu/vous	*tu*	15.0	98/652
	vous	26.9	154/572
Topic	serious	25.4	93/366
	not serious	18.5	159/858
Quoted speech	formal	41.7	5/12
	not quoted	20.4	242/1184
	not formal	17.9	5/28
Interlocutor L1	L1 French	21.9	139/635
	L2 French	19.2	113/589

Age	younger	16.5	54/328
	older	22.1	198/896
Length of residence	> 10 years	19.9	193/970
	< 10 years	23.2	59/254
Gender	male (*n* = 1)	50.7	34/67
	female (*n* = 9)	18.8	218/1157
Conversation portion	0-5 min	22.2	22/99
	5-10 min	20.4	44/216
	10+ min	20.5	186/909
TOTAL		20.6%	252/1224

Table G-5. Non-significant factor groups for Near-NSs in Pau and Lille combined (ne-retention); cf. Table 4-25

Factor Group	**Factor**	**% ne-retention**	**N**
Clause type	main clause	16.6	292/1764
	subordinate clause	30.0	89/297
Reinforcing adverb	adverb	19.0	22/116
	no adverb	18.5	359/1945
Negator	*pas*	17.8	307/1729
	other	22.3	74/332
Quoted speech	formal	35.3	6/17
	not quoted	18.5	367/1988
	not formal	14.3	8/56
Tu/vous	*tu*	15.2	227/1489
	vous	26.9	154/572
Interlocutor L1	L1 French	19.2	218/1136
	L2 French	17.6	163/925
Age	younger	15.0	128/855
	older	21.0	253/1206
Length of residence	> 10 years	19.4	248/1280
	< 10 years	17.0	133/781
Gender	male (*n* = 5)	24.0	121/504
	female (*n* = 14)	16.7	260/1557
Conversation portion	0-5 min	17.9	38/212
	5-10 min	17.0	52/306
	10+ min	18.9	291/1543
Site	Pau	20.6	252/1224
	Lille	15.4	129/837
TOTAL		18.5%	381/2061

Table G-6. Non-significant factor groups for NSs in Lille (ne-retention); cf. Table 4-26

Factor Group	Factor	% ne-retention	N
Object clitic	clitic	6.6	5/76
	no clitic	10.8	60/558
Reinforcing adverb	adverb	3.3	2/61
	no adverb	11.0	63/573
Lexicalization	not lexicalized	12.5	58/464
	lexicalized	4.1	7/170
Negator	*pas*	10.2	54/529
	other	10.5	11/105
Topic	serious	13.3	24/180
	not serious	9.0	41/454
Quoted speech	formal	27.3	3/11
	not quoted	19.7	61/310
	not formal	7.7	1/13
Conversation portion	0-5 min	4.9	4/81
	5-10 min	10.1	9/89
	10+ min	11.2	52/464
TOTAL		10.3%	65/634

Table G-7. Non-significant factor groups for NSs in Pau (ne-retention); cf. Table 4-27

Factor Group	Factor	% ne-retention	N
Reinforcing adverb	adverb	9.9	8/81
	no adverb	6.2	75/1208
Verb type	auxiliary	10.6	13/123
	main	6.0	44/733
	copula	6.2	20/323
	modal	5.5	6/110
Quoted speech	formal	33.3	2/6
	not quoted	6.5	80/1229
	not formal	0.0	0/54
Conversation portion	0-5 min	6.3	10/158
	5-10 min	4.7	7/150
	10+ min	6.7	66/981
Speaker guise	English identity	5.9	36/606
	French identity	6.9	47/683
TOTAL		6.4%	83/1289

Table G-8. Non-significant factor groups for NSs in Lille and Pau combined (ne-retention); cf. Table 4-28

Factor Group	Factor	% ne-retention	N
Object clitic	clitic	8.0	13/163
	no clitic	7.6	133/1760
Reinforcing adverb	adverb	6.9	10/144
	no adverb	7.8	138/1779
Lexicalization	not lexicalized	9.9	131/1327
	lexicalized	2.9	17/596
Negator	*pas*	7.3	118/1622
	other	10.0	30/301
Gender	male (*n* = 3)	9.9	67/677
	female (*n* = 2)	6.5	81/1246
Conversation portion	0-5 min	5.9	14/239
	5-10 min	6.7	16/239
	10+ min	8.2	118/1445
TOTAL		7.7%	148/1923

Table G-9. Non-significant factor groups for near-native interlocutors in Lille (ne-retention); cf. Table 4-29

Factor Group	Factor	% ne-retention	N
Relative clause	relative clause	27.2	3/11
	no relative clause	12.3	38/310
Reinforcing adverb	adverb	20.0	2/10
	no adverb	12.5	39/311
Verb type	auxiliary	19.0	8/42
	main	13.2	25/189
	copula	10.0	6/60
	modal	6.7	2/30
Lexicalization	not lexicalized	14.6	35/239
	lexicalized	7.3	6/82
Negator	*pas*	9.5	33/349
	other	11.1	8/72
Topic	serious	13.6	15/110
	not serious	12.3	26/211
Quoted speech	formal	40.0	2/5
	not quoted	12.3	38/308
	not formal	12.5	1/8
Interlocutor ID	1L	16.7	7/42
	2L	14.0	6/43
	3L	10.0	3/30
	4L	15.4	4/26
	5L	7.3	3/14

Interlocutor ID	6L	9.8	5/51
	7L	0.0	0/15
	8L	20.9	9/43
	9L	13.3	4/30
Conversation portion	0-5 min	7.0	3/43
	5-10 min	10.5	4/38
	10+ min	14.2	34/240
TOTAL		12.8%	41/321

G.2. Non-significant factor groups for subject doubling

Table G-10. Non-significant factor groups for SA learners (subject doubling); cf. Table 5-22

Factor Group	**Factor**	**% SD**	**N**
Subject type	proper noun	36.0	28/75
	common noun	24.8	70/282
	indefinite/other pronoun	0.0	0/12
Subject specificity	specific	32.2	87/270
	non-specific	9.1	4/44
	generalizing	14.0	8/57
Subject animacy	animate	8.3	10/120
	inanimate & material	36.1	13/36
	inanimate &non-material	33.6	47/140
	place	38.7	29/75
Relative clause	relative clause	29.4	5/17
	no relative clause	26.3	93/341
Negation	affirmative	26.5	91/344
	ne-deletion	72.7	8/11
	ne-retention	0.0	0/16
Preverbal clitics	none	28.4	98/345
	reflexive clitic	0.0	0/4
	object clitic	4.5	1/22
Tu/vous	*tu*	26.1	69/264
	vous	37.5	21/56
	unknown	17.6	9/51
Conversation portion	0-5 min	33.3	23/69
	5-10 min	25.8	24/93
	10+ min	24.9	52/209
Gender	Male (n = 5)	26.6	68/256
	Female (n = 2)	31.2	31/115
Interlocutor L1	L1 French	32.1	34/106
	L2 French	24.5	65/265
TOTAL		26.4%	98/371

Table G-11. Non-significant factor groups for SA interlocutors (subject doubling); cf. Table 5-24

Factor Group	Factor	% SD	N
Clause type	matrix	68.6	83/121
	subordinate	55.6	10/18
	relative	0.0	0/1
	si	0.0	0/1
Verb type	copula	75.6	62/82
	transitive	62.1	18/29
	intransitive	50.0	11/22
	modal	28.6	2/7
	passive	--	--
Other pre-verbal material	adverb	66.7	2/3
	feedback	100.0	1/1
	hesitation	77.8	7/9
	none	64.4	67/104
	parenthetical	75.0	3/4
	prepositional phrase	68.4	13/19
Negation	affirmative	64.9	85/131
	ne-deletion	88.9	8/9
	ne-retention	0.0	0/1
Subject definiteness	definite	66.4	87/131
	indefinite	60.0	3/5
	quantified	60.0	3/5
TOTAL		66.0%	93/141

Table G-12. Non-significant factor groups for Near-NSs in Lille (subject doubling); cf. Table 5-25

Factor Group	Factor	% SD	N
Other pre-verbal material	adverb	60.0	3/5
	feedback	0.0	0/1
	hesitation	57.9	11/19
	multiple elements	55.6	5/9
	none	52.4	164/313
	parenthetical	88.9	8/9
	prepositional phrase	48.1	13/27
Subject specificity	specific	60.6	168/277
	non-specific	30.8	16/52
	generalizing	37.0	20/54
Relative clause	relative clause	85.0	17/20
	no relative clause	51.5	187/363

Subject definiteness	definite	55.0	179/326
	indefinite	52.6	10/19
	quantified	20.7	6/29
	verb (not applicable)	100.0	9/9
Negation	affirmative	52.8	180/341
	ne-deletion	67.7	21/31
	ne-retention	27.3	3/11
Preverbal clitics	none	53.3	196/368
	reflexive clitic	30.0	3/10
	object clitic	33.3	5/15
Conversation portion	0-5 min	65.0	26/40
	5-10 min	53.3	32/60
	10+ min	51.6	146/283
Gender	Male (*n* = 4)	51.9	94/181
	Female (*n* = 5)	54.5	110/202
Interlocutor L1	French	58.1	136/234
	English	45.6	68/149
Age	younger	52.5	127/242
	older	54.6	77/141
Length of residence	shorter (< 10 years)	52.5	127/242
	longer (> 10 years)	54.6	77/141
TOTAL		53.3%	204/383

Table G-13. Non-significant factor groups for Near-NSs in Pau (subject doubling); cf. Table 5-26

Factor Group	**Factor**	**% SD**	**N**
Subject type	proper noun	36.8	28/76
	common noun	24.9	70/281
	indefinite/other pronoun	0.0	0/12
Subject definiteness	definite	42.3	243/575
	indefinite	45.0	18/40
	quantified	22.2	12/54
	verb (not applicable)	87.5	14/16
Preverbal clitics	none	42.7	263/616
	reflexive clitic	41.9	13/31
	object clitic	28.9	11/38
Tu/vous	*tu*	47.2	171/362
	vous	35.9	116/323
Conversation portion	0-5 min	38.5	30/78
	5-10 min	38.7	36/93
	10+ min	43.0	221/514
Gender	Male (*n* = 1)	35.0	21/60
	Female (*n* = 9)	42.6	266/625

Age	younger	52.8	105/199
	older	37.4	182/486
Length of residence	shorter (< 10 years)	52.5	127/242
	longer (> 10 years)	54.6	77/141
Interlocutor identity	English identity	41.8	142/340
	French identity	42.0	145/345
TOTAL		41.9%	287/685

Table G-14. Non-significant factor groups in Near-NSs for Pau and Lille combined (subject doubling); cf. Table 5-28

Factor Group	**Factor**	**% SD**	**N**
Preverbal clitics	none	47.1	459/974
	reflexive clitic	39.0	16/41
	object clitic	30.2	16/53
Other pre-verbal material	adverb	57.1	12/21
	feedback	50.0	1/2
	hesitation	40.4	21/52
	multiple elements	52.4	11/21
	none	45.3	391/863
	parenthetical	63.3	19/30
	verb (not applicable)	45.6	36/79
Subject specificity	specific	47.5	396/751
	non-specific	44.4	47/149
	generalizing	28.6	48/168
Conversation portion	0-5 min	47.5	56/118
	5-10 min	44.4	68/153
	10+ min	46.0	367/797
Tu/vous	*tu*	50.3	375/745
	vous	35.9	116/323
Interlocutor L1/identity	English	42.9	210/489
	French	48.5	281/579
Site	Pau	41.9	287/685
	Lille	53.3	204/383
TOTAL		46.0%	491/1068

Table G-15. Non-significant factor groups for NS interlocutors in Lille (subject doubling); cf. Table 5-30

Factor Group	Factor	% SD	N
Relative clause	relative clause	83.3	15/18
	no relative clause	54.4	136/250
Verb type	copula	67.1	98/146
	transitive	45.1	23/51
	intransitive	35.1	13/37
	modal	50.0	9/18
	passive	50.0	8/16
Subject definiteness	definite	56.1	129/230
	indefinite	58.3	7/12
	quantified	31.3	5/16
	verb (not applicable)	100.0	10/10
Subject specificity	specific	64.0	121/189
	non-specific	40.0	20/50
	generalizing	34.5	10/29
Subject animacy	animate	45.7	53/116
	inanimate & immaterial	58.3	7/12
	inanimate & material	62.1	64/103
	place	73.0	27/37
Conversation portion	0-5 min	50.0	10/20
	5-10 min	63.7	28/44
	10+ min	55.4	113/204
TOTAL		56.3%	151/268

Table G-16. Non-significant factor groups for NS interlocutors in Pau (subject doubling); cf. Table 5-31

Factor Group	Factor	% SD	N
Verb type	copula	68.9	268/389
	transitive	63.4	137/216
	intransitive	54.1	99/183
	modal	50.0	26/52
	passive	60.0	9/15
Subject definiteness	definite	63.2	494/782
	indefinite	60.9	14/23
	quantified	44.1	15/34
	verb (not applicable)	100.0	16/16
Subject specificity	specific	67.0	458/684
	non-specific	47.3	44/93
	generalizing	47.4	37/78

Preverbal clitics	none	64.4	508/789
	reflexive clitic	44.4	8/18
	object clitic	47.9	23/48
Conversation portion	0-5 min	65.1	71/109
	5-10 min	70.2	73/104
	10+ min	61.5	395/642
Tu/vous	*tu*	61.3	152/248
	vous	63.8	387/607
TOTAL		63.0%	539/855

Table G-17. Non-significant factor groups for NS interlocutors in Lille and Pau (subject doubling); cf. Table 5-32

Factor Group	**Factor**	**% SD**	**N**
Verb type	copula	68.4	366/535
	transitive	59.9	160/267
	intransitive	50.9	112/220
	modal	50.0	35/70
	passive	54.8	17/31
Subject definiteness	definite	61.6	623/1012
	indefinite	60.0	21/35
	quantified	40.0	20/50
	verb (not applicable)	100.0	26/26
Conversation portion	0-5 min	62.8	81/129
	5-10 min	68.2	101/148
	10+ min	60.0	508/846
Tu/vous	*tu*	58.7	303/516
	vous	63.8	387/607
Site	Pau	63.0	539/855
	Lille	56.3	151/268
TOTAL		61.4%	690/1123

Table G-18. Non-significant factor groups for near-native interlocutors in Lille (subject doubling); cf. Table 5-34

Factor Group	**Factor**	**% SD**	**N**
Clause type	matrix	58.2	92/158
	subordinate	42.3	22/52
	relative	66.7	2/3
	si	0.0	0/1
Verb type	copula	67.7	67/99
	transitive	48.2	27/56
	intransitive	34.8	16/46
	modal	54.5	6/11
	passive	33.3	1/3
Subject definiteness	definite	52.4	99/189
	indefinite	66.7	2/3
	quantified	46.2	6/13
	verb (not applicable)	100.0	10/10
Subject specificity	specific	53.7	87/162
	non-specific	53.8	14/26
	generalizing	59.3	16/27
Subject animacy	animate	42.7	50/117
	inanimate & material	75.0	3/4
	inanimate & immaterial	70.4	50/71
	place	60.9	14/23
Preverbal clitics	none	56.2	113/201
	reflexive clitic	66.7	2/3
	object clitic	18.2	2/11
Negation	*ne*-deletion	71.4	5/7
	affirmative	54.2	109/201
	ne-retention	42.9	3/7
Conversation portion	0-5 min	61.9	13/21
	5-10 min	54.8	17/31
	10+ min	53.4	87/163
TOTAL		54.4%	117/215

CURRICULUM VITAE

MARK ALAN BLACK

EDUCATION

Ph.D., French with a concentration in French Linguistics. February 2021. *Interlocutor Effects on Sociolinguistic Variation in L2 French*. Indiana University, Bloomington, IN. Minors: Linguistics; Second Language Studies

M.A., French. May 2012. Indiana University, Bloomington, IN

***Licence* degree, Linguistics** (mention *bien*). June 2010. Université de Rouen, Rouen, France. Specialization in teaching French as a foreign language

M.A., French. May 2006. University of Wyoming, Laramie, WY

B.A., French (*cum laude*). May 2003. University of Wyoming, Laramie, WY.
Minor: Information Management

PROFESSIONAL EXPERIENCE

Lecturer of French. Aug 2020 – present. Washington State University, Pullman, WA

Visiting Lecturer & Course Supervisor, French. Aug 2016 – Oct 2017; Aug 2019 – May 2020. Indiana University, Bloomington, IN

Lecturer of French. Aug 2017 – May 2019. Depauw University, Greencastle, IN

Maître de langue / Lecteur d'anglais (English instructor / lecturer). Sep 2014 – Aug 2016. Université de Pau et des Pays de l'Adour, Pau, France

Associate Instructor. July 2011 – July 2014. Indiana University, Bloomington, IN

Financial Coordinator / Grammar Instructor. June-July 2012, 2013 in Brest, France. Indiana University Honors Program in Foreign Languages, Bloomington, IN

Lecteur d'anglais. Nov 2009 – June 2010. Université de Rouen, France

Assistant étranger (TAPIF). Oct. 2008 – April 2009. Le Havre, France

French Teacher. Dec 2007 – June 2008. Laramie High School, Laramie, WY

Graduate Assistant. Aug 2004 – May 2006. University of Wyoming, Laramie, WY

Assistant étranger (TAPIF). Oct 2002 – April 2004. Le Blanc, France & Orléans, France

LANGUAGES

Native speaker of English
Near-native proficiency in French
Intermediate proficiency in Spanish
Low-Intermediate proficiency in German

PUBLICATIONS

Black, M. (submitted). Implications of grammatical gender in measures of L2 sociolinguistic competence: The case of subject doubling in French. In D. Ayoun (Ed.), *The Acquisition of grammatical gender.* Amsterdam: John Benjamins.

Dekydtspotter, L., Black, M., Frimu, R. & Panwitz, A. (2018). Animacy-based processing loads in anaphora resolution in (non-native) French: Evidence for privileged interface representations. *Meaning and structure in second language acquisition: In honor of Roumyana Slabakova (*pp. 95 – 122). *Amsterdam:* John Benjamins.

Black, M., Frimu, R., Panwitz, A., Dekydtspotter, L. & Lorente-Lapole, A. (2013). Computational cycles in (second) language processing: Cyclic versus non-cyclic integration in French. In W. Orman & M. J. Valleau (Eds.), *Proceedings of the 38th Annual Boston University Conference on Language Development* (pp. 51 – 63). Somerville, MA: Cascadilla Press.

CONFERENCE PRESENTATIONS

"Interlocutor effects on sociolinguistic variation in L2 French: Symmetry in the system?" *38th Second Language Research Forum*, Michigan State University, Sep. 22, 2019.

"Sociolinguistic understanding: The impact of the interlocutor on variation in L2 French." *Understanding and Misunderstanding,* French and Italian Graduate Student Conference, Indiana University, March 1, 2019.

"Subject doubling in advanced and near-native speakers of French." *New Ways of Analyzing Variation (NWAV) 47*, New York University, Oct. 21, 2018.

"Interlocutor language ability as a factor on L2 sociolinguistic variables: The case of null objects in L2 French." New Interdepartmental Conference on Linguistics Area Studies (NICOLAS), Indiana University, April 13, 2018.

"Interlocutor native language influence on L2 sociolinguistic patterns of French *ne*-deletion." *36th Second Language Research Forum*, The Ohio State University, Oct. 14, 2017.

"Interlocutor language ability as a factor on L2 sociolinguistic variables: The case of L2 French ne-deletion." *Crossing Borders, Breaking Walls*, French and Italian Graduate Student Conference, Indiana University, March 10, 2017.

"Interlocutor L1 effects on L2 sociolinguistic competence: A variationist analysis of L2 French ne-deletion." *Symposium: Interlocutor Individual Differences*, Indiana University, Oct. 1, 2015.

"Variable interrogative structures in French L2-L2 informal discourse." *Variation in Language Acquisition 2*, Université de Grenoble III, Dec. 4, 2014.

"Levels of structure in anaphoric processes in L1 and L2 French. Co-authored paper presented by L. Dekydtspotter and R. Frimu, *33rd Second Language Research Forum*, University of South Carolina, Oct. 25, 2014.

"Informal discourse markers in native-nonnative speaker interactions: Evidence of distinct adherence to target language norms." PRISMES workshop, *Learning a second language: spoken interaction between native and non-native speakers*, Sorbonne Nouvelle Paris III, June 21, 2014.

"Computational cycles in (second) language processing: Cyclic versus non-cyclic integration in French." Co-authored paper presented by L. Dekydtspotter and R. Frimu, *38th Annual Boston University Conference on Language Development*, Boston University, Nov. 2, 2013.

"A modifier-complement asymmetry along cyclic-movement chains in L1-English L2-French sentence processing." Co-authored paper presented with A. Lorente-Lapole and A. Panwitz. *32nd Second Language Research Forum*, Brigham Young University, Nov. 1, 2013.

"Translating the ungrammatical: Making sense of things that don't make sense." *21st Graduate Student Conference, French and Romance Philology*, Columbia University, March 2, 2012.

DEPARTMENTAL PRESENTATIONS

"Interlocutor effects on sociolinguistic variation in L2 French: Symmetry in the system?" Second Language Studies Colloquium, Indiana University, Sep. 13, 2019.

"L'interrogation variable dans le discours informel de l'apprenant du français langue seconde." Seminar series in Linguistics and Didactics, Université de Pau et des Pays de l'Adour (Pau, France), Nov. 24, 2014.

"Computational cycles in (second) language processing: cyclic versus non-cyclic integration in French." Co-authored paper presented with L. Dekydtspotter, R. Frimu and A. Panwitz at Second Language Studies Colloquium, Indiana University, Sep. 27, 2013.

"Translating the ungrammatical: Making sense of things that don't make sense." Department of French and Italian Student-Faculty Forum, Indiana University, Feb. 22, 2012.

GUEST LECTURES

"Sociolinguistic competence in L2 French." Guest lecture in Julie Auger's course F672 Dialectologie et sociolinguistique françaises, Indiana University, Sep. 28, 2017.

"Sociolinguistic competence in L2 French: How to design a study and measure it?" Guest lecture in Julie Auger's course L441 Field Methods in Sociolinguistics, Indiana University, Sep. 21, 2017.

"Computational cycles in (second) language processing: cyclic versus non-cyclic integration in French." Guest lecture with R. Frimu and A. Panwitz in Rex Sprouse's course SLST 301 Introduction to L2 Acquisition, Indiana University, Oct. 28, 2013.

TRAVEL AND RESEARCH GRANTS

Grace P. Young Fellowship, $300. Travel grant from Department of French and Italian, Indiana University, for conference presentation at the Symposium on Interlocutor Individual Differences (AILA Research Network), Indiana University, Oct. 2014.

Grace P. Young Fellowship, $350. Travel grant from Department of French and Italian, Indiana University, for conference presentation at 32nd Second Language Research Forum, Brigham Young University, Oct. 2013.

Grace P. Young Fellowship, $350. Travel grant from Department of French and Italian, Indiana University, for conference presentation at Graduate Student Conference Department of French and Romance Philology, Columbia University, March 2012.

Albert Valdman Graduate Student Research Travel Fund, $800. Travel grant for dissertation research in France, May 2013 and November 2014.

SERVICE & EXTRA-CURRICULAR ACTIVITIES

Chapter reviewer for *The Routledge Handbook of Second Language Acquisition and Sociolinguistics* (Eds. S. M. Gass & A. Mackey), forthcoming.
Committee member, Humanities and Social Sciences, Washington State University, 2020
French Faculty liaison, Advance College Project (ACP) at Indiana University, 2019-2020
Volunteer, World Language Festival, Indiana University, Oct. 2018
Panel Moderator, *Generative Approaches to Language Acquisition North America (GALANA) 8* conference, Indiana University, Sep. 2018
Volunteer, French kiosk at Lotus Blossoms World Bazaar (youth education day), April 2018
Volunteer, French Club, Indiana University, 2010-2014, 2016-2017, 2019-2020
Co-organizer, Fest Noz Breton Dancing night, University of Georgia, Nov. 2017
Co-organizer, TPRS demonstration for teachers of refugee students, Université de Lille 3, France, 2016
Organizer, English-French trivia night for students, Université de Pau, France, 2014-2016

Panel moderator, Department of French and Italian Graduate Student Conference, Indiana University, March 2013
Volunteer, *New Ways of Analyzing Variation (NWAV) 41* conference, Indiana University Bloomington, Nov. 2012
Organizer, French Conversation Table, University of Wyoming, 2005-2007

AWARDS

Grace P. Young Graduate Award ($500), Departmental award for outstanding achievement in French Graduate Studies, Indiana University, 2015.
Marvin D. Moody Fellowship, Indiana University, 2010-2011
Siren Memorial Scholarship, University of Wyoming, 1999
Trustees Superior Student Scholarship, University of Wyoming, 1998-2002
Member of American Association of Teachers of French, Pi Kappa Alpha Fraternity, Golden Key, Phi Kappa Phi and Phi Beta Kappa

UNIVERSITY TEACHING EXPERIENCE

Washington State University (Fall 2020 – present)
French 101 First Semester French
French 120 French Culture
French 203 Third Semester French
French 305 Intermediate Conversation II
French 307 Intermediate Speaking and Listening

Indiana University (Fall 2010 – Summer 2014; Fall 2016 – Spring 2018; Fall 2019 – Spring 2020. Supervisor, Fall 2016 – Spring 2017; Fall 2019 – Spring 2020)
F100 Elementary French I (face-to-face and online)
F150 Elementary French II
F200 Second Year French I
F250 Second Year French II (face-to-face and online)
F316 Conversational Practice
F491 Elementary French for Graduate Students (non-French majors)
F492 Reading French for Graduate Students (non-French majors)

Depauw University (Fall 2017 – Spring 2019)
FREN 102 Elementary French II
FREN 110 Review of Elementary French (accelerated course)

Université de Pau et des Pays de l'Adour (Fall 2014 – Spring 2016)
Langue orale (English) 1st year: Phonology lab, comprehension lab, conversation class
Langue orale (English) 2nd year: Phonology lab, phonology lecture
Langue orale (English) 3rd year: Comprehension and expression

Métiers de l'enseignement, de l'éducation et de la formation (MEEF) 1st year: Oral and comprehension skills lecture

Université de Rouen (Fall 2009 – Spring 2010)
MEEF 1st and 2nd year: Oral skills practice

University of Wyoming (Fall 2004 – Spring 2008)
French 1010 1st Year French I (in-person and distance learning)
French 1020 1st Year French II (in-person and distance learning)
French 2030 2nd year French I (distance learning)